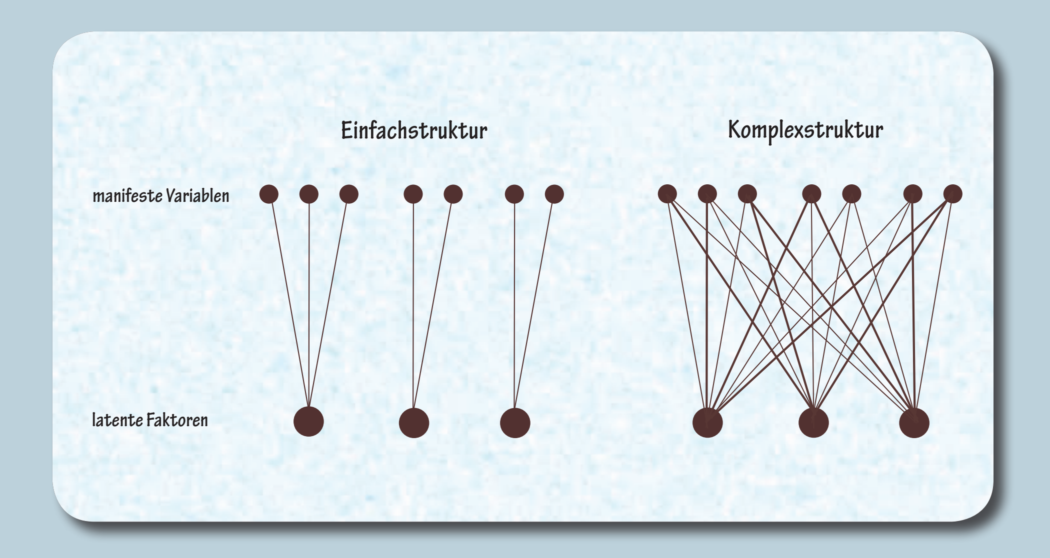
Suitbert Ertel

Komplexität modellieren

Faktorenanalyse am Scheideweg



Universitätsverlag Göttingen

Suitbert Ertel

Komplexität modellieren



erschienen im Universitätsverlag Göttingen 2011

Suitbert Ertel

Komplexität modellieren

Faktorenanalyse am Scheideweg



Universitätsverlag Göttingen
2011

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Anschrift des Autors
Suitbert Ertel
Tobias-Mayer-Weg 3
37077 Göttingen

e-mail: sertel@uni-goettingen.de

Satz und Layout: Suitbert Ertel; Franziska Lorenz
Umschlaggestaltung: Franziska Lorenz

http://univerlag.uni-goettingen.de
ISBN: 978-3-86395-013-2

Inhalt

Kapitel 2 Komplexe Strukturen aufspüren.
Faktorenanalysen mit Variminrotation

Kapitel 3 Komplexität intellektueller Leistungen.
Grundintelligenz (g) und Lernkapital (l).

Verwendete Abkürzungen

AV	Abhängige Variable
CFA	Konfirmatorische Faktorenanalyse
CFT	Culture-fair Intelligence Test
CS	Komplexstruktur
CSM	Komplexstruktur-Modellierung
EFA	Exploratorische Faktorenanalyse
ERS	Empirisches Relationssystem
FA	Faktorenanalyse
FRS	Formales Relationssystem
FRT	Figure Reasoning Test
g	Generalfaktor
I-S-T	Intelligenz-Strukturtest
MDS	Multidimensionale Skalierung
MTMM	Multitrait-Multimethod-(Analyse)
NMDS	Nichtmetrische multidimensionale Skalierung
PCA	Principal Component Analysis (Hauptkomponenten-Analyse)
SD	Social desirability
SEM	Structural equation modeling (Strukturgleichungsmodellierung)
SS	Simple Structure (Einfachstruktur)
SSM	Simple structure modeling (Modellierung als Einfachstruktur)
UV	Unabhängige Variable

Vorwort

Paul Barrett

Stellen Sie sich vor, Sie hätten sich als Psychologe oder Psychologin mit Ihren modernen quantitativen Methoden in eine multivariate kausale Theorie vertieft, um psychologische Merkmale und Phänomene zu durchleuchten. Die Verwendung einer konfirmatorischen Faktorenanalyse (CFA) und das Modellieren von Beziehungen zwischen Ihren Daten wäre Ihre nächste Obliegenheit. Ein unzeitgemäßes Vorgehen, das gelegentlich noch bei älteren Fachvertretern anzutreffen ist, bestünde darin, die exploratorische Faktorenanalyse (EFA) mit Hauptkomponentenanalyse und orthogonalen Rotationen einzusetzen. Der Titel des Buches, das Sie in den Händen halten, „Faktorenanalyse am Scheideweg", lässt Sie vielleicht denken, dass die antiquierte EFA hier endgültig zu Grabe getragen werden soll. Tatsächlich verfolgt dieses Buch aber das entgegengesetzte Ziel: Es will die EFA zu neuem Leben erwecken. Ist das denn möglich - werden Sie fragen. Was für einen Sinn hätte ein solches Vorhaben, wäre es zu begrüßen?

Das Konzept der Einfachstruktur wird in diesem Buch problematisiert. Es hat seit den Vierziger Jahren des letzten Jahrhunderts die Entwicklung von Taxonomien der Persönlichkeitseigenschaften und kognitiven Fähigkeiten wesentlich mit geprägt. Zu den Ergebnissen zählen die Big Five, das HEXACO Persönlichkeitsinventar, die 16PF von Cattell, Eysencks Superfaktoren, um nur einige „Simple-Structure"–Produkte zu nennen. Die Literatur über kognitive Fähigkeiten quillt über mit solchen Modellen, die einen Höhepunkt mit dem Meisterwerk von John Carroll von 1993 erreichten, in dem Ergebnisse von Faktorenanalysen mit Datenmatrizen für 468 Fähigkeiten berichtet wurden. Die Modelle der Einfachstruktur waren für die meisten quantitativ arbeitenden Psychologen, die Theorien zur „Struktur" der menschlichen Persönlichkeit und der Intelligenz entwickelten, grundlegend und richtungweisend.

Von Psychometrikern mit einer zeitgemäßeren Ausbildung wird die Simple Structure Doktrin allerdings durchaus noch als ein notwendiges Ingredienz ihrer Verfahren betrachtet, sie wird routinemäßig bei der Konstruktion konfirmatorischer Modelle verwendet. Wenn Items zu einem vom Forscher konzipierten Faktor eindrucksmäßig nichts beizutragen scheinen, setzt er ihre Gewichte auf Null, die

Software führt die entsprechenden Befehle automatisch aus. Mit Modellbildungen dieser Art werden die Einfachstrukturen Thurstones festgeschrieben.

Doch weicht die Realität, die man modellieren will, von der Simple Structure Doktrin ständig ab, was die Modellautoren auch immer wieder zugeben. Deshalb versagen ihre Modelle so oft nicht nur bei präzisen Fit-Tests, sondern auch bei toleranteren Prüfungen. Die früheren EFA-Forscher wussten sehr wohl, dass die meisten psychologischen Variablen, insbesondere solche aus Persönlichkeitsuntersuchungen, bei den Faktoren, an denen sie interessiert waren, die erwünschten Null-Ladungen meist nicht aufwiesen. Man half sich mit einer Art Daumenregel: „Eine Ladung größer als .03 bei einem Nebenfaktor invalidiert die betreffende Variable als Indikator eines Hauptfaktors".

Im Grunde hätten diese regelmäßig auftretenden Abweichungen der beobachteten von erwarteten und erwünschten Ergebnissen zu der Erkenntnis führen können, dass die Grundbedingungen menschlichen Verhaltens keine mathematisch abbildbaren Größen darstellen wie Temperatur, Masse, Zeit usw. Um diesem misslichen Problem zu begegnen, sind in jüngster Zeit Strukturgleichungsansätze entwickelt worden (ESEM: Marsh et al. , 2010), die mithilfe von Zugeständnissen einige Vorteile der früheren Forschung zurück gewinnen wollen. Dem Forscher wird jetzt erlaubt, die Arbeitsweise eines EFA-Anwenders zum Teil wieder aufzunehmen, ohne dass er deswegen auf die neueren Modell-prüfenden CFA-Methoden zu verzichten hätte.

Indessen, ist es nicht auch vorstellbar, dass eine fundamentalere Revision der exploratorischen Faktorenanalyse versucht werden könnte, die mehr zuwege bringt als geringfügige Korrekturen, wie sie durch ein Flickwerk der CFA erreicht werden? Als selbstkritische Wissenschaftler sollten wir immer daran denken, dass die Struktur, die wir mit einem Modell der Realität gewinnen, von der Methodologie abhängt, die wir dafür einsetzen. Michael Maraun z. B. hat 1997 gezeigt, dass bei einer gleichen Datenlage die scheinbar unveränderliche Fünffaktorenstruktur der Persönlichkeit durch ein geeignetes Vorgehen in eine Zweifaktorenstruktur (Radex-Struktur) überführt werden kann. Die Wahl eines statistischen Modells zur Repräsentation von Realität kann statistik-intern und mathematisch getroffen werden. Das Simple Structure-Modell ist so entstanden. Man kann sich aber auch zunächst am jeweiligen Gegenstand orientieren und an dem, was man über ihn schon weiß, um erst dann das formale Abbild der Phänomene den Phänomenen selbst anzupassen.

Hier kommt die Message von Suitbert Ertel voll zum Tragen. Ganz offensichtlich hat er als Fachpsychologe dem Gegenstand seiner Wissenschaft das erste Wort gelassen und sich darum bemüht, dem adaptiv-integrierenden Geschehen, das für uns bei kognitiven Operationen höchste Transparenz erreicht, mithilfe der forma-

len Modelle, die die Faktorenanalyse entwickelt hat, eine optimale Repräsentanz zu ermöglichen.

Warum haben ihm die bestehenden Faktorenmodelle nicht gereicht? Das Streben Thurstones nach formaler Einfachheit war von einer fehlorientierten, historisch damals verständlichen Erwartung einer „sozialen Physik“ getragen. Man war der Meinung, dass die Psychologie mit ihren erforschbaren Funktionen als ein Sondergebiet der Naturwissenschaft, die reduktionistisch operieren will, zu behandeln sei. Strukturen hatten so einfach wie möglich zu sein, der damaligen Generation von Psychologen war die Vorstellung von grundlegender Komplexität und von interaktiv-adaptiven Systemen fremd. Mit dem Hinzukommen neuer Erkenntnisse aus der experimentellen Psychologie, aus den Neurowissenschaften und anderen Disziplinen, die gezwungen waren, Systeme mit Komponenten in Wechselwirkung zu erforschen und entsprechende Modelle zu entwickeln, wissen wir, dass die kausalen Prozesse, die dem menschlichen Verhalten zugrunde liegen, biologisch komplexer Natur sind und der Selbstorganisation dienen. Bei solchen Einsichten im 21. Jahrhundert wirkt eine Konzeption psychologischer Traits mit „Einfachstruktur“ eher rückständig und befremdend.

Das Neue an Suitberts Auffassung ist, dass er die Einfachstruktur als Modell für menschliche Verhaltensweisen und Einstellungen gänzlich verwirft. Stattdessen verwendet er eine Methodologie, durch welche Kovarianzstrukturen dazu gebracht werden, den multiplen Quellen der Varianz gerecht zu werden, die auf die einzelnen Variablen eingewirkt haben. Seine Methodologie wird nicht nur psychologischen Fragestellungen gerecht. Sie ist auch bei anderen biologischen und bei physikalischen Messvariablen anwendbar, wie im zweiten Kapitel seines Buches nachzulesen ist. Eine atemberaubende Wende im Umgang mit multivariaten Daten findet hier statt, in der Tat. Sie widerspricht allen CFA-Konzeptionen und einer Generation altvertrauter, eher nur tolerierter als befriedigender Simple Structure-basierter Taxonomien.

Mit zahlreichen empirischen Beispielen in seiner Monographie zeigt Suitbert, dass sich nach Anwendung seines Varimin-Verfahrens die Interpretation der resultierenden Faktoren sehr unterscheidet von dem, was man mit Anwendung der EFA gewöhnlich zustande bringt. Man kommt ins Staunen, wenn man sieht, wie sehr sich die Deutung der Varimin-Faktoren mit den Vorstellungen deckt, die man sich a priori macht über die Vorgänge in kognitiven Systemen. Während Simple Structure-Modelle für ihre Glaubwürdigkeit ein nur ziemlich pauschales Konzept diskreter Einheiten erfordern, gelangt man mit Varimin-Faktoren zu einem Einblick in die Dynamik integrativ operierender Wechselwirkungen innerhalb der Systeme. Mit Varimin bekommt man bei der Durchführung von Faktorenanalysen endlich das Steuer in die Hand. Mit diesem Verfahren kann man nach Problemlösungen suchen, die mit der Realität der jeweils untersuchten Systeme korrespondieren. In

der Tat ist Varimin ein nicht nur einfaches, sondern auch sehr adaptives Verfahren.

Was kann ein moderner mit multivariaten Daten quantitativ arbeitender Psychologe aus alledem machen? Suitbert greift Fragen auf, die die übergeordneten Ziele der multivariaten statistischen Analyse zum Thema haben. Will man weiterhin eine an der Einfachstruktur orientierte Beschreibung der Daten erreichen, will man die latenten Variablen, die man durch manifeste Variablen von vorne herein repräsentiert sehen möchte, weiterhin mit kausalen Funktionen verknüpfen? Oder sollte man den Varimin-Ansatz bevorzugen, bei dem latente Variablen zunächst zu erschließen sind, bevor sie mit ihren kausalen Wechselwirkungen modelliert werden?

In beiden Fällen werden latente Variablen mit manifesten in Beziehung gesetzt. Doch mit Simple Structure will man Einflüsse fragwürdig rotierter Faktoren auf ganze Datensätze ansetzen. Man deklariert zu diesem Zweck Cluster der Variablen zu Repräsentanten der Faktoren, den einzelnen Variablen wird dabei keine mehrfaktorielle Basis zugesprochen. Im Varimin-Modell dagegen repräsentieren Variablen nicht einzelne Faktoren, eine mehrfaktorielle Basis wird den Variablen belassen. Die Bedeutung der Faktoren wird hier nicht durch Variablencluster repräsentiert, sie wird aus den mehrfaktoriellen Ladungsverteilungen bei den Variablen abgeleitet. Hat man die Bedeutung der Faktoren gefunden, wird die Kovarianz in der Variablenmatrix als Ergebnis eines kausalen Zusammenwirkens dieser Faktoren verständlich.

An dieser Stelle sollte man auch den Begriff der Ursache durchdenken. Ursachen wirken bei Datensätzen unserer Forschung kaum jemals in Isolation, wie es das Simple Structure-Modell in Analogie zu elementaren physikalischen Größen voraussetzt. So kann man sich die Variablen Länge, Masse, oder Zeit und andere aus dem internationalen Einheitensystem der Physik (SI) durch additive Operationen entstanden denken. Eine größere Länge lässt sich durch das Zusammenfügen kleinerer Längen herstellen. Obgleich die Länge vieler Objekte durch Temperatur veränderbar ist, ist doch weder die Länge von Temperatur noch die Temperatur von Länge abhängig. Für solche elementaren Messgrößen werden Simple Structure-Regressionen nicht gebraucht.

Setzt man nun für Phänomendomänen der Psychologie die Faktorenanalyse ein, dann sollte man nicht wie Thurstone und Nachfolger damit rechnen, SI-analoge elementare Dimensionen wie Länge oder Temperatur zu finden. Die in Faktoranalysen unserer Domäne verwendeten Variablen sind nicht elementar, sie entstehen durch Interaktion anderer Variablen. Simple Structure-Lösungen bei solchen Daten setzen sich über die Komplexität der manifesten Variablen blind hinweg, sie führen zu Fehlrepräsentationen der Zusammenhangsverhältnisse. Varimin kann man als eine Methodologie des „derived measurement" verstehen,

Beobachtungsvarianzen bei manifesten Variablen werden als eine Funktion von zwei oder mehr latenten Variablen aufgefasst, die den manifesten Messungen zugrunde liegen. Sogar in der Physik wird man mit komplexerem Geschehen konfrontiert. Dichte z. B. ist eine Funktion der Masse (gm) und des Volumens (Millimeter oder Kubikzentimeter), Watt ist eine Funktion von Länge, Masse und Zeit usw. Was Suitbert mit seiner Methodologie gezeigt hat, ist dass die traditionelle EFA, CFA und SEM zu ganz anderen Repräsentationen der Realität führen als die, die sich durch Varimin ergeben, und er sagt auch, warum nicht beide Methodologien richtig sein können.

Ein einzelnes Buch kann kaum ein ganzes Gebiet wissenschaftlicher Anstrengungen mit ihren historisch geronnen Folgen verändern. Aber es kann wichtige, bislang übersehene Fragen aufwerfen. Suitbert hat dies getan und darüber hinaus hat er Varimin für Analysen vieler Datensätze ausgetestet. Das mag genug sein für einen Kickstart für Master- und Doktorarbeiten, mit denen komplexe Datenstrukturen anstelle von Einfachstrukturen zu erforschen wären. Sie als Leserin/Leser werden wahrscheinlich kaum daran vorbei kommen, neugierig zu werden auf Auswirkungen der Varimin-Forschung auf zukünftige theoretische Entwicklungen in Ihrem Fachgebiet.

Auch wenn Sie skeptisch bleiben sollten: Wir haben hier mit einer faszinierenden methodologischen Kehrtwendung zu tun. Die neue Methodologie könnte auf Dauer neue Wege zur Erforschung psychischen Geschehens überhaupt erschließen, lassen Sie sich das mal durch den Kopf gehen. Ein kleiner Vorschlag zum Abschluss: Befreien Sie sich von den Zwängen der Simple Structure, dann wird Ihnen Suitbert eine Pandora-Büchse voller Forschungsmöglichkeiten überlassen, in der Sie, wenn Sie sie aufmachen, wie in der griechischen Mythodologie vorerst vielleicht nur Würmer sehen. Doch wenn Sie neugierig bleiben und sie ein zweites Mal öffnen, könnten Sie von Forschungsergebnissen überrascht werden, die Sie schon lange erhofft haben.

Für mich war es ein Privileg, dieses Vorwort zu schreiben mit dem Gedanken, dass die Leser von den neuen Möglichkeiten für ihre Forschung ausgiebig und Gewinn bringend Gebrauch machen.

* Die Übersetzung aus dem Englischen besorgte Elisabeth Cott.

Literatur

Carroll, J.B. (1993) *Human Cognitive Abilities: a Survey of Factor Analytic Studies*. New York: Cambridge University Press.

Maraun, M.D. (1997) Appearance and Reality: Is the Big Five the Structure of Trait Descriptors? *Personality and Individual Differences*, 22, 5, 629-647.

Marsh, H., Ludtke, O., Muthen, B., Asparouhov, T., Morin, A.J.S., Trautwein, U., & Nagengast, B. (2010) A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22, 3, 471-491.

Paul Barrett

Honorary Professor of Psychology, University of Auckland, NZ
Adjunct Associate Professor of Psychology, University of Canterbury, NZ
CEO, Advanced Projects R&D Ltd., NZ
Email: paul@pbarrett.net
Web: www.pbarrett.net

Einführend

„Rebel with a cause" – so betitelte Hans-Jürgen Eysenck seine Autobiographie. Ihm, dem friedlichen Anwalt wissenschaftlicher Vernunft, war in Auseinandersetzung mit Irrtümern des herrschenden Zeitgeistes eine kämpferische Rolle zugefallen. Ich möchte mit der vorliegenden Monographie vier Kapitel vorlegen, bei deren Abfassung ich mich - wohl oder übel - auch auf eine Auseinandersetzung mit herrschenden Irrtümern einzulassen hatte. Auch in meinem Fall könnte der Eindruck eines rebellischen Protestes dem Leser das friedliche Motiv der Vernunft verdecken, das allein mich dabei leitete.

Am Anfang war ich ein begeisterter, aber dennoch unzufriedener Anwender der Faktorenanalyse. Warum hatten sich schon die Pioniere dieser Methodologie (Spearman, Cattell, Burt, Thurstone) in grundsätzlichen Fragen zerstritten? Warum führten die Verfahren, mit denen man dimensional einfache Strukturen finden wollte, im Ergebnis zu einem Sammelsurium zusammenhangloser Konstrukte? Irgendetwas konnte nicht stimmen, das war mein Verdacht.

Durfte ich aber als bloßer Anwender der angesehenen mathematischen Verfahren an dem zweifeln, was die Experten für unumstößlich hielten? Konnten Generationen hochkarätiger Methodologen geirrt und ihren Irrtum nicht bemerkt haben?

Ich hatte den Verdacht, dass es nicht richtig war, die Faktorenmatrix, die aus einer Interkorrelationsmatrix der untersuchten Variablen extrahiert wird, dem Prinzip der ‚simple structure' auszuliefern, das Thurstone 1935/47 eingeführt hatte. Thurstone hatte die komplexe Initiallösung einer Faktorenanalyse für nicht interpretierbar gehalten, weil in einer Initiallösung singuläre Variablen durchgängig bei mehreren Faktoren Ladungsgewichte zeigten. Er hielt es für unmöglich, die Bedeutung eines Faktors aus mehrfaktoriell gewichteten Variablen abzuleiten. So transformierte er die originären Faktorkoordinaten durch geometrische Rotation so lange, bis er für einzelne Variablen möglichst nur einen Faktor mit möglichst großem Gewicht gewonnen hatte, während sich für diese Variablen bei den übrigen Faktoren Gewichte nahe Null ergaben. Trotz vieler verbleibender Fragen im Detail, welche Nachfolger und Kritiker Thurstones kontrovers beantworteten, erschien allen sein Prinzip der Einfachstruktur eine notwendige Richtschnur zu geben. Doch meinte ich überall in der Natur Vielfachbedingtheit, Komplexität, nicht Simplizität. Mit dem Ideal der Einfachstruktur schien mir das atomistische Denken fortzudauern, das in der Frühgeschichte der Psychologie oft dominierte. Eine holistische, für komplexe Strukturen und Gestaltbildungen offene Herange-

hensweise an psychologische Phänomene schien mir auch für die Faktorenanalyse unverzichtbar. Aber ich war mir nicht sicher.

Erst nach rund 30 Jahren Inkubation eines unorthodoxen selbstunsicheren Denkens besann ich, das Prinzip Thurstones radikal umzustoßen und sein Ideal der Einfachstruktur gegen das der Komplexstruktur auszutauschen. Bei der Suche nach erforderlicher Evidenz wurde ich durch ein neues von mir entworfenes Computerprogramm unterstützt. Es rotiert Faktoren in der Weise, dass untersuchte Variablen faktorielle Komplexstrukturen manifestieren. Ich nannte das Verfahren „Variminrotation“, weil sein Ziel dem der Einfachstruktur, das mit der konventionellen Varimaxrotation erreicht werden soll, entgegen gesetzt war. [1] Als ich dann feststellte, dass die Komplexstrukturen der Variablen, die mit Varimin gewonnen wurden, sich sehr viel leichter und besser interpretieren ließen als die Einfachstrukturen, die sich mit Varimax ergaben, fingen für mich die Würfel an zu fallen.

Ich möchte mich an dieser Stelle bei Wegbegleitern bedanken, die mich bei meinem gewagten Unternehmen durch Interesse, Rat und Vertrauen unterstützten. Dazu gehören:
Paul Barrett, Jürgen Bredenkamp, Elisabeth Cott, Edgar Erdfelder, Hans-J. Eysenck, Sibyl Eysenk, Herbert Götzl, Jürgen Guthke, Gerd Lüer, Matthias Romppel, Pierre Sachse, Tatiana Schnell, Rolf Steyer, Michaela Strack und Hans-Werner Wendt.

[1] Für die Programmierarbeit bin ich Dr. Uwe Engeland dankbar, der das Programm unter meiner Leitung ausarbeitete. Das Programm minimiert die Spaltenvarianz der Faktorladungen, während diese durch Varimax maximiert wird. Matthias Romppel überließ mir freundlicherweise sein Computerprogramm greomat.exe, mit dem die visuellen Darstellungen von Kapitel 2 (Faktorenladungen) herstellbar waren.

Kapitel 1

Kritik am Dogma „Simple Structure"

„Dinge zu bezweifeln, die ganz ohne weitere Untersuchung
jetzt geglaubt werden, das ist die Hauptsache überall"
Georg Christoph Lichtenberg (1742-1799).

Vorspann zu Kapitel 1

Im ersten Kapitel bemühe ich mich, dem auf den Grund zu gehen, was ich für einen Irrtum der Faktorenanalytiker in der Thurstone-Tradition halte, nämlich die Rotation extrahierter Faktoren nach dem Prinzip Simple Structure (SS) (Thurstone, 1935,1947). Thurstones Ideal der SS wurde in der Praxis nie erreicht, statistische Hilfsmaßnahmen, die man sich im Laufe der Zeit ausdachte, führten bestenfalls zu partiellen Korrekturen. Meine Untersuchung führt zu der Einsicht, dass SS nur scheinbar eine zwingende Antwort ist auf die Frage, auf welche faktorielle Struktur hin die Rotation der primär extrahierten Faktoren ausgerichtet werden sollte.

Hat es nicht Kritiker gegeben, die den Irrtum des Einfachstruktur-Ideals erkannten oder zumindest erahnten? Es hat einige gegeben. Warum blieben ihre Stimmen ungehört? War die spätere Wende weg vom exploratorischen Forschen, das so viele fragwürdige Ergebnisse lieferte, hin zum konfirmatorischen Forschen und den anschließenden Strukturgleichungsmodellen (SEM) ein Fortschritt? Kaum, denn am Prinzip der Einfachstruktur (Simple Structure Modelling, SSM) haben die

konfirmatorischen Verfahren festgehalten. Man ließ das Prinzip unangetastet. SSM hat alle Varianten des Modellierens latenter Faktoren geprägt, der exploratorisch ermittelten und der mit konfirmatorischer Absicht eingeführten Faktoren, ohne dass die Konsequenzen der SSM erkannt wurden. Wie konnte es dazu kommen?

Dass ich in den folgenden Kapiteln mit meiner Kritik an SSM überzeugen möchte, mag Lesern missfallen, die an der Gültigkeit von SSM nie gezweifelt haben. Auch bin ich mir bewusst, dass man mit theoretischen Überlegungen, die ich im ersten Kapitel ausführe, widersprechende Dauerüberzeugungen kaum verändern kann. Die Zuversicht beim Vortrag meiner Kritik sollte mit der Lektüre der Kapitel zwei bis vier verständlicher werden, in denen ich darlege, was mit der Faktorenanalyse gewonnen wird, die sich am Scheideweg der Forschung vom Zwang des SSM befreit und die Aufdeckung komplexer Varianzquellen von Variablen zum Ziel eines wirklichkeitsnäheren Modellierens macht.

Die ersten drei Beiträge sind 2009 und 2010 im derzeit noch wenig bekannten *Journal Psychologie des Alltagshandelns* veröffentlicht und für die vorliegende Monographie überarbeitet worden. Der vierte Beitrag wurde mit erforderlichen Kürzungen in *Personality and Individual Differences* in englischer Sprache veröffentlicht. Auf zwei weitere Beiträge, darunter ein umfasssender über Persönlichkeit (Ertel, a), wurde wegen Überlänge und mit der Vornahme, ihn an anderer Stelle ohne Kürzung zu veröffentlichen, hier verzichtet. Die vier Beiträge des vorliegenden Readers bauen als Kapitel einer Monographie aufeinander auf mit dem Vorteil, dass sie unabhängig voneinander verständlich sind. Es kommt nur zu wenigen Überschneidungen, bei denen der Grundgedanke aus sich ergänzenden Perspektiven dargestellt wird. Am Ende der Monographie finden sich Abstracts der vier Kapitel. Die ‚Vorspann'-Texte zu Anfang der Kapitel verbinden diese zu einem Projekt mit gleichem Ziel.

Ausgangslage und Zielsetzung

Ich glaube, dass durch die Rotation der Faktoren nach dem Prinzip Simple Structure (SS) empirische Variablendomänen generell inadäquat modelliert werden. Das Thurstone-Prinzip gibt der Bedingungskomplexität, die bei den untersuchten Variablen in der Regel vorliegt, keine Chance, sich zu manifestieren. Alternative Modelle werden nicht erwogen.

Im Folgenden gebe ich nicht nur meine eigenen Ansichten wieder, sondern – meist mit wörtlichen Zitaten – auch die Ansichten kritischer Autoren, die ich als Wegbereiter ausfindig machen konnte, deren Stimmen aber von der Majorität der

Fachvertreter ignoriert werden. Eine Alternative zur konventionellen SSM, eine Transformation zur Komplexstruktur (Complex Structure Modelling, CSM), wird in Kapitel 2 entworfen und dort mit zehn Anwendungsbeispielen erläutert. Im ersten Kapitel werden offensichtliche Schwächen und Fehler des bisherigen methodischen Denkens diskutiert. Man hätte sich diese ersparen können, wenn man die Analysetechniken von Anfang an auf die Gegebenheiten und Sachverhalte, auf die sie angewendet werden, sinnvoll bezogen und dem gesunden Menschenverstand etwas Mitspracherecht eingeräumt hätte, für den die Bedingungskomplexität des Erlebens und Verhaltens (und aller seiner Variablen) selbstverständlich ist.

01. Zur Lage der faktorenanalytischen Forschung

Düstere Titel und Metaphern enttäuschter Autoren sind ein Zeichen dafür, dass die Faktorenanalyse unter chronischen Anomalien leidet: *Unbehagen in der Faktorenanalyse* (Kallina, 1967); *erschreckende Unverbindlichkeit*, *Dilemma und Aufbaukrise* (Orlik, 1967); *Subjectivity in factor analysis* (Horn, 1967); *Zwiespältigkeit faktorieller Untersuchungen* (Meili, 1968); *Destruction of generality* (Davies, 1971); *Zufallsprodukte und Scheinrealisationen* (Greif, 1972); *Unsinniger Aufwand* (Revenstorf, 1978); *Rotationsambiguität von Faktorladungen* (Buse & Pawlik, 1978); *Dubious legacy* (Schönemann, 1981); *Faktoranalytis* (Jäger & Hörmann, 1981); *Mythos Faktorenanalyse* (Lenk, 1983); *Factors are fictions*, Revelle (1983); *Morass of factor analysis* (Eysenck, 1992); *Psychopathology of factor indeterminacy* (Schönemann, 1996) [2]; *Pathology of psychometrics* (Borsboom, 2003).

---

[2] Schönemann (1981) und Steiger (Steiger & Schönemann, 1975) gehören nach Guttman (1955) zur mittleren Generation der Methodiker, die an der Faktorenanalyse Kritik übten, ihre Kritik war scharf („theoretical problems", „users are generally uninformed about the defects of this model", p. 175, 188). Doch lag ihr Akzent nicht auf *Simple Structure*, sondern auf der formalen Unentscheidbarkeit über die Geltung faktorieller Dimensionen (‚indeterminacy') und deren Nicht-Identifizierbarkeit (‚lack of identifiability'). Schönemann & Steiger (1976) entwickelten nach Reanalysen von dreizehn publizierten faktorenanalytischen Studien, die zur vernichtenden Kritik an diesen führte (Schönemann & Wang, 1972), ein alternatives Verfahren der multivariaten Datenreduktion (*Regression Component Decomposition*, *RCD*), das im Vergleich zur Faktorenanalyse größere begriffliche Klarheit und rechnerische Effizienz sowie die Möglichkeit einer Modell-Falsifizierung versprach. Doch blieb dieser Ansatz angesichts des Siegeszugs der ‚multiplen Faktorenanalyse' Thurstones unbeachtet. Zur Frage der Transformation der durch RCD ermittelten ‚Komponenten' bot der alternative Ansatz tatsächlich nur wenig Neues. Zudem waren die Ergebnisse der RCD von denen der Faktorenanalyse Thurstones offenbar nicht sehr verschieden.

Die Mängel der Methode werden von der Mehrheit der Experten bagatellisiert und gern mit Optimismus überdeckt: Es gebe doch ‚bedeutsame Ergebnisse', meint Pawlik (1977) in einem umfassenden deutschsprachigen Überblick über die ersten Jahrzehnte der faktorenanalytischen Forschung.

Aus der Sicht unkonventionell eingestellter Autoren, die der Misere ins Auge sehen, ist die faktorenanalytische Forschung fehl gelaufen: *„Exploratory factor analysis has never been developed to anything approaching its full promise and potential, despite the eighty year history of its efforts…"* (Yates, 1987, p. 325). Blinkhorn (1997) behandelt in einem Überblick über „fifty years of test theory" die Faktorenanalyse und stellt fest, dass auch die „considerable technical strides" der letzten Jahrzehnte und die „well-known contributions of Jöreskog and McDonalds" die Misere im Grunde nicht verändert haben: *„How curious…that we are so little further forward in our understanding of the psychology of individual differences as a result of these advances… Can anyone identify a single publication in the last 50 years in which the use of factor analysis has led to counter-intuitive, or surprising, or genuinely enlightening outcomes? …"* (Blinkhorn, 1997, p. 181). Schon vor 50 Jahren hätte man aufmerken können bei dem, was Schönemann (1994) retrospektiv über Louis Guttman berichtete, der 1955 eine ‚Grabrede' [eulogy] über die multiple Faktorenanalyse hielt: *„It was left to Louis Guttman to read the eulogy (p. 209, p. 406): The era of Multiple Factor Analysis had come to an end - for knowledgeable people at any rate. … It was logical, then, to ask: What lies ahead for Factor Analysis? (Guttman, 1958). He answered it with a vision that challenged habits of thought that had lead nowhere. This vision he kept pursuing for the rest of his life"*, Schönemann & Borg (1996, p. 249) ziehen Bilanz: *„Heute weiß man, dass die von Thurstone eingeläutete Ära der ‚explorativen Faktorenanalyse' wenige Einsichten von Bestand gebracht hat"*. [3]
Zwei Kalamitäten der EFA sind zu beklagen:

- *EFA-Forschung hat zu einer Unzahl von Konstrukten in der Psychologie geführt und somit das Gegenteil von dem erreicht, was man mit ihr erreichen wollte.*

---

[3] Übersichtsreferate über die exploratorisch-faktorenanalytische Forschungsmethodik mit Anwendungsorientierung aus jüngerer Zeit – die meisten ohne grundlegende Kritik – stammen von Moosbrugger & Hartig (2002) und Fabrigar et al. (1999), denen Artikel mit ähnlichem Ziel vorausgingen: Stevenson (1993), Tinsley & Tinsley (1987), Ford et al. (1986), Glass & Taylor (1966), Cattell (1965), Peel (1953). Die am häufigsten zitierten englisch-sprachigen Lehrbücher zur exploratorischen Faktorenanalyse sind – mit z. T. Angabe der letzten und ersten Auflage: Comrey & Lee (1992, 1973), Child (1990,1971), Gorsuch (1983, 1974), Harman (1976, 1968), Mulaik (1972). Deutschsprachige Äquivalente sind Weber (1978), Revenstorf (1976), Überla (1977, 1971). Eine deutschsprachige Einführung als Lehrbuch-Anhang bietet Lienert (1969) und als Lehrbuch-Kapitel (Nr. 15) Bortz (2004, 1977).

Die EFA hat zum Ziel, die Vielzahl miteinander korrelierter manifester Variablen durch eine geringere Zahl faktorieller Variablen, denen die Rolle *latenter Dimensionen* zugesprochen wird, sparsam und theoretisch vorteilhaft zu beschreiben.[4] Nun hat sich in den Jahrzehnten der EFA-Forschung herausgestellt, dass die Produktion faktorieller Variablen, die den beobachtbaren Variablen zugrunde liegen sollen, unerschöpflich ist. Lee Sechrest wies auf der 11th European Conference on Personality (2002) in Jena hin auf die Schwemme von Konstrukt-Variablen in der Psychologie, die er nach einem Autor, der sie hochrechnete, auf 7,800 bezifferte, viele davon sind Neuschöpfungen exploratorischer Faktorenanalysen.

Ist Sechrest's Schätzung unrealistisch? Nach einem Internet-Download von Artikel-Titeln mit dem Stichwort „Scale" aus der Datenbank PubMed fanden sich immerhin 889 unterschiedliche Skalen-Bezeichnungen, von „***A**bel and Becker Cognition Scale*" bis „***Z**ung Self Rating Anxiety Scale*". Die meisten Skalen sind faktorenanalytisch erarbeitet oder bearbeitet worden, so dass man bei einer geschätzten Faktorenzahl von 3 bis 4 pro Skala etwa 3,000 faktorielle Konstrukte ansetzen kann, die bisher allein im klinisch-psychologischen und medizinischen Raum entstanden sind. Hinzuzufügen sind die in der nicht-klinischen differentiellen Psychologie entwickelten Skalen und die aus ihnen gewonnenen Konstrukte, von denen allein die Zeitschrift *Personality and Individual Differences* mit jedem Heft so viel Neues bietet, dass der oben genannte Schätzwert realistisch erscheint. *„The idle practice of producing new personality scales continues unabated, making it less likely that they will ever arrive in the promised land of the paradigm which alone would endow our efforts with scientific respectability"* (Eysenck, 1992, p. 672). [5]

---

[4] Mit dem Begriff ‚Dimension' wird der Anspruch einer Metrik angemeldet, die nicht hinterfragt wurde. So wie die drei Dimensionen des euklidischen Raumes der Ortung von Objekten dienen, so meinte man, die Hauptvariablen psychologischer Beobachtung mithilfe von faktoriell gewonnenen ‚Dimensionen' verorten zu können. Dabei wird die Metapher der Raumdimensionen überstrapaziert (z.B. auch durch den Ausdruck ‚Semantic Space' von C. E. Osgood). Der Terminus Dimension, der im mathematischen Sprachgebrauch berechtigt ist, ist zur Bezeichnung von Varianzquellen auf der Objektebene irreführend und entbehrlich.

[5] Die Gesamtheit typologischer Konstruktionen der differentiellen Psychologie bis 1970, nicht nur faktorenanalytischer Provenienz, hat Ruttkowski (1974) zu erfassen versucht. Dazu kommentiert Sponsel (1998): *„Weltweit gibt es über 1000 Persönlichkeits-, Charakter- oder Wesenstypologien (Ruttkowski, 1974). Die Mehrzahl davon dürfte … umstritten sein. Viele sind in der Kultur- und Wissenschaftsgeschichte un-*

- *Den faktoriellen Konstrukten fehlen schon innerhalb der Analysen, die sie hervorbringen, theoretisch verwertbare Verknüpfungen.*

Die faktoriellen Konstrukte aus ein und derselben Analyse sind Unikate und als solche voneinander isoliert. Die Produktion solcher Unikate führt zu bloßen Aggregaten. Erst durch Beziehungen untereinander könnten Begriffe, die bei äußerlich bleibender Betrachtung isoliert erscheinen, sich in eine Ordnung fügen. Doch *„…Simple structure may have resulted in an analysis into too many unrelated, and UNRELATABLE, primaries [primary factors]"* (so Stephenson schon 1956, p. 6, Großbuchstaben durch SE). Der derzeit umfassendste kritisch-historische Überblick über die persönlichkeitspsychologische EFA-Forschung von Andresen (1998) vermittelt den Eindruck einer chaotischen Ergebnisvielfalt.

Hat das ‚Big Five' Faktoren-Modell, das in der Persönlichkeitsforschung entstand, dort stürmisch begrüßt wurde und hohes Ansehen erreichte,[6] den theoretischen Mangel beseitigt? Man hat sich damit begnügt, in der „chaotic plethora of personality constructs" (Funder, 2001, p. 200) fünf leidlich invariante Konstrukte herauszustellen. Doch ist Invarianz von Konstrukten allein kein hinreichendes Kriterium für ihre Güte, auch fehlerhafte Konstrukte können ‚robust' sein. Weit überzogen ist der Optimismus einiger Big-Five Proponenten, die die Rolle ihrer Faktoren in Analogie setzen zur Rolle der chemischen Elemente (begonnen hat damit offenbar Goldberg, 1981), deren Entdeckung im 19. Jahrhundert eine naturwissenschaftliche Revolution auslöste.[7] So wie sich aus chemischen Elementen

---

tergegangen. Viele überschneiden sich. Es hat den Anschein, als ob beliebig viele Konstruktionen möglich sind - je nach unterschiedlichen Zielen und Zwecken." Ähnlich urteilen Gigerenzer & Strube (1987, p. 85): *„Es ist die Crux der faktorenanalytischen Forschung, so viele ‚anerkannte' Persönlichkeitsfaktoren aufgefunden zu haben, dass selbst bei einfacher Dichotomisierung der Dimensionen die Zahl der dadurch zur Verfügung stehenden Quadranten höherer Ordnung mit schätzungsweise 2^{50} … etwa vierhunderttausendmal so groß ist wie die derzeitige Weltbevölkerung."*

6 Einen Eindruck vom Elan der Big-Five-Bewegung verschafft eine umfassende Sammlung von Zitaten von Chris Brand. Sie ist bildschirmzugänglich über die URL http://www.cycad.com/site/Brand/quotes/q03.html

7 Blinkhorn (1997, p.180) kritisiert überzogene Hoffnungen schon bei den Pionieren der Faktorenanalyse: *"The words they use, for example 'primary mental abilities' (Thurstone) or 'source traits' (R.B. Cattell), are witness to the faith and trust placed in factor analysis as revealing the psychological analogue of the periodic table of elements, or the list of subatomic particles."*

Moleküle zusammensetzen, sollten sich komplexere differentielle Konstrukte der Psychologie aus fünf elementareren Konstrukten zusammen fügen lassen. Dies glaubten z. B. einige Optimisten, die die scheinbar elementaren „großen Fünf" durch kombinatorische Verwendung in drei Superkonstrukte überführen wollten (‚Resilience', ‚Overcontrol' und ‚Undercontrol' , s. Caspi, 1998, Robins et al., 1998, Asendorpf & van Aken, 1999). Die Meinung hat sich verbreitet, dass sich die zukünftige Faktorenproduktion im Persönlichkeitsbereich durch Korrelationen mit den Big Five zu legitimieren habe.[8]

Doch „trägt dieser Vergleich [mit den chemischen Elementen] nicht weit" (Lukesch & Kleiter, 1974, p. 294). Denn H, He, C, Ca, N usw. haben einen funktional definierbaren Platz innerhalb des Periodensystems. Dass sich aus Atomen bestimmte Moleküle bilden, ist den atomaren Bindungseigenschaften zu verdanken, die sich aus der Zahl der Protonen in den Atomkernen, aus der Elektronendichte usw. ergeben. Aufgrund gleichartiger Bestandteile ihres Aufbaus stehen chemische Elemente miteinander in Beziehung. Bei den Big Five-‚Dimensionen' der Persönlichkeit aber sind keine gleichartigen Bestandteile erkennbar, die zu einem Ähnlichkeits- oder Verschiedenheitsurteil berechtigten. Nur wenige kritische Beobachter nehmen daran Anstoß, z. B. Briggs, 1989, und Block, 1995 [9]. Die „großen Fünf" sind wie alle anderen Konstrukt-Produkte der Faktorenanalyse nicht miteinander verknüpfbar, eine Analogie zu den Elementen der Chemiker Dimitri Mendelejev (1834-1907) oder Lothar Meyer (1830-1895) ist verfehlt. Eher könnte man die Faktoren der EFA den vermeintlich unteilbaren Atomen Heraklits (B.C. ca. 540-480) gleich setzen, die ein ‚atomistisches' Denken festschrieben, das

[8] Ozer & Reise 1994 *"characterized the big five as the 'latitude and longitude' along which any new personality construct should be routinely mapped"* (Funder, 2001, p. 200).

[9] Briggs (1989): *"… a coherent and falsifiable explanation for the five factors has yet to be put forward". "There is no theoretical reason why it should be these five rather than some other five"* (p. 249). *"The structure of trait attributions may not correspond straightforwardly to the deep structure or neurophysiological basis of human tendencies."* (p. 250). *"Perhaps the critical step in elucidating these concepts [interpreting the five factors] … is the specification of their exact nature.: What are the elements or components of each factor? How are they interrelated?"* (p. 253). Block (1995) zitiert Briggs und bemängelt noch präziser*: "No functioning psychological 'system', with its rules and bounds, is designated or implied by the 'Big Five' formulation; it does not offer a sense of what goes on within the structured, motivation-processing, system-maintaining individual"* (p. 188).

erst mit den Entdeckungen der atomaren und subatomaren Strukturen aufzugeben war. [10]

An dieser Stelle wäre auch auf H. J. Eysenck einzugehen, der auf die atheoretischen Auswüchse der EFA-Forschung sensibel reagierte. Den Big Five-Forschern warf er vor, Faktorenanalysen mit lediglich taxonomischer Abgrenzungsabsicht zu betreiben und dabei an der psychometrischen Oberfläche zu kleben, anstatt biologisch-kausal interpretierbare Zusammenhangsmodelle zu entwickeln (Eysenck, 1992,1997). Eysencks eigener Versuch wäre anzuführen, der eine Verbindung unter den drei Dimensionen seines PEN-Modells (**P**sychoticism, **E**xtraversion, **N**euroticism) zu erreichen suchte, indem er für sie eine differentielle Hirngrundlage postulierte und den Dimensionen anatomische Strukturen und neuro-psychologische Funktionen zuwies, die zusammenwirken sollten. So begrüßenswert auch Eysencks biopsychologisches Ziel grundsätzlich war - er hätte zunächst die Mängel der Methode untersuchen sollen.[11] Die eigentliche Ursache der

[10] Atomistisches Denken hat der Psychologie auch sonst geschadet – das war die Klage der Gestalt-Psychologen (s. Metzger, 1975), die offenbar wenig Wirkung zeigte, wenn man nicht gerade wie der Verfasser dieses Artikels vom relationalen Denkstil dieser Richtung noch beeinflusst wurde: Die vorliegende Kritik am Atomismus der Simple Structure wäre ohne Wolfgang Metzgers Grundbegriffe vom *Zusammenhang,* vom *Bezugssystem*, der *Zentrierungsdynamik* usw. (Metzger, 1941) vermutlich nicht entstanden. Kritik am Atomismus der Thurstone-Methodologie wurde schon von Richard Meili (1969) formuliert, der noch vor Ort in der „Berliner Schule“ der Gestaltpsychologie geprägt wurde (durch Wolfgang Köhler und Kurt Lewin, Meili war Promovend in Berlin, 1924), und der sich dem Zwang der SSM nur mit ständigem Widerspruch fügte: “Das Ökonomieprinzip der Minimalstruktur“ habe zwar „einen gewissen heuristischen Wert“, sei aber „für die Natur nicht bindend“. Die Annahme, man könne Denken und Intelligenz theoretisch „auf die einfachste Weise erfassen“, sei fraglich. „*Wie wäre die Chemie wohl mit der analogen Hypothese gefahren, dass in jedem vorgefundenen Stoff möglichst nur ein Element vorkommen sollte*“ (p. 288). Auch Jäger & Hörmann (1981, p. 418) haben einzuwenden: „*Die nach wie vor verbreitete und perpetuierte Analogie zur Chemie, nach der Faktoren quasi wie chemische Elemente , 'gesucht', 'entdeckt' und als 'existierend' bestätigt werden, ist irreführend.*“

[11] *“How should the Big-5-or-6 be understood in psychological terms? Sadly, despite many years of research - especially into extraversion - the picture is still very unclear (see e.g. A. Gale & M. W. Eysenck, 1992, Handbook of Individual Differences: Biological Perspectives; G. Matthews, 1993, in A. Smith & D. Jones, Factors Affecting Human Performance.) Here are some possibilities that still look viable yet falsifiable.”* (C. R. Brand, Quelle s. Fußnote 5) .

von ihm beklagten „morass of factor analysis“ (1992, p. 672) hat er nicht erkannt, sie liegt nicht dort, wo er sie vermutete.

Faktorenanalytische Datenverarbeitung findet man auch in zahlreichen nichtpsychologischen Disziplinen (vgl. Rummel, 1970, Reymert & Jöreskog, 1993, und Abbildung 1.01), Unbehagen wird nicht selten auch dort bekundet. So hört man z. B. aus den Reihen der Geowissenschaftler, die sich methodisch der SSM-orientierten EFA bedienten, etwa bei Davies (1971): *„[By varimax] we may be butchering our results; cutting up the body of generality into a set of unrelated fragments without ever realizing that these fragments can ever be considered as part of a larger entity“* (p. 113). Davies bezeichnet die Wirkung der Varimax-Rotation auf faktorisierte Datensätze wiederholt als „zerstörerisch“. Dies wird in Abschnitt 02 näher beleuchtet.

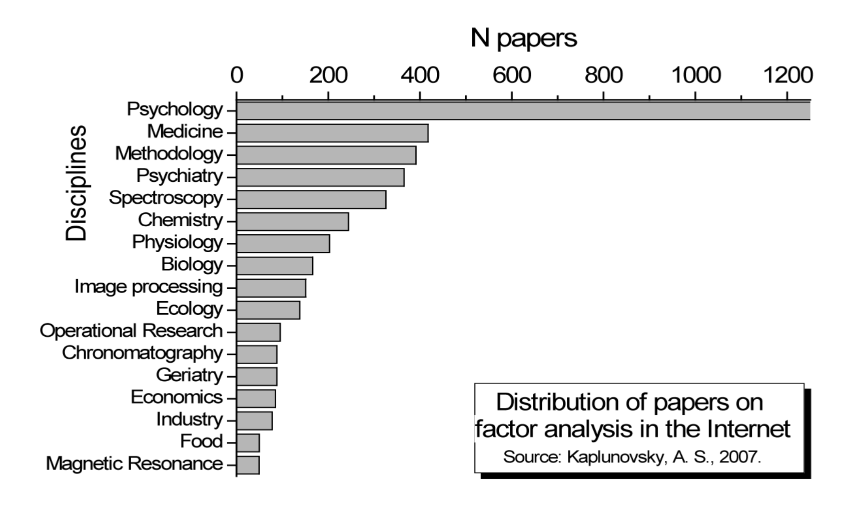


Abbildung 1.01
Anzahl faktorenanalytischer Arbeiten, differenziert nach Disziplinen, ermittelt von A. S. Kaplunowsky (2007) aus Internetquellen

02. Das Prinzip der Einfachstruktur als Dogma

Der Bericht über die beklagenswerte Lage der faktorenanalytischen Forschung hatte zum Ziel, die Wurzeln der Kalamität zu verstehen. Als Ursache für das verbreitete „Unbehagen in der Faktorenanalyse“ hat man bislang meist subjektive Willkür bei der Auswahl von Verfahrensweisen angesehen, so etwa bei der Auswahl der zu korrelierenden Variablen, bei der Auswahl der zu testenden Personen-

stichproben, bei der Entscheidung über die Zahl der zu extrahierenden Faktoren, bei der Entscheidung über schief- oder geradwinklige Rotation, bei der dann immer noch anstehenden Auswahl aus einer Vielzahl von Auswertungsmöglichkeiten (s. Finch & West 1997, p. 464 ff). Doch halte ich die Wirkungen dieser Entscheidungswillkür auf die Endergebnisse der Faktoranalysen für geringfügig gegenüber dem, was die Misere in der EFA-Forschung tatsächlich hervorruft. Die Wurzeln der Kalamität sind, wie angekündigt, das SSM [12]

Der Grundgedanke der Einfachstruktur ist weithin bekannt, er ist „intuitively compelling" (Kaiser, 1958, p. 188). Im Koordinatensystem der initialen Faktoren ergeben sich Häufungen von Variablenpunkten (s. Abbildung 1.02A), die den Statistiker veranlassen, durch die Punktwolken Regressionslinien zu ziehen, so dass das Quadrat der Abstände der Punkte von der Linie ein Minimum erreicht. Mit diesem Ziel werden die Koordinaten des Systems so gedreht, dass diese mit dem Regressionsvektor zusammen fallen (Abbildung 1.02B). Da nacheinander immer zwei Koordinaten gleichzeitig gedreht werden, wird bei einer Drehung, wenn mehrere Punktwolken in der Ebene vorhanden sind, eine gleichzeitige Annäherung beider Koordinaten optimiert. Die konsequenteren unter den SSM-Methodikern übertreffen dabei die rechtwinklig rotierenden Forscher an Radikalität, indem sie Schiefwinkligkeit in Kauf nehmen, nur um eine maximale Annähe rung an eine SS zu erreichen.[13]

---

[12] Schon Velicer (1977) fand, dass die Extraktionsverfahren maximum likelihood, image analysis und principal components analysis bei neun Test-Datensätzen im Ergebnis 'extremely similar' (p.18) ausfielen. Goldberg (1990) konnte bei einem repräsentativen Datensatz von Eigenschaftsbeschreibungen das Big-Five-Faktorenmodell fast invariant reproduzieren, egal ob er die Methode der Faktorenextraktion variierte (principal components, principal factors, alpha-factoring, image-factoring, maximum-likelihood) und egal ob er orthogonal (Varimax) oder schiefwinklig (Oblimin) rotierte. Doch diese methodischen Variationen wurden ausnahmslos mit Simple Structure-Orientierung vorgenommen. Alternativen zur Simple Structure wurden nicht diskutiert, obgleich zumindest Goldberg nach seinem Eysenck-Zitat (1981) eine Abkehr von Simple Structure als Denkmöglichkeit in Erwägung zog: *"Correlational psychology cannot in the nature of things come up with objective, universally agreed dimensions or categories; there are innumerable, mathematically equivalent ways of rotating factors, for instance, and no statistical magic key (not even simple structure) can close the door on alternative solutions (p. 43)."*

[13] Warburton (1963) beschreibt in einer Übersicht über die analytischen Methoden der Faktorenrotation die Logik mathematischer Sparsamkeit mit einfachen Worten: *"He [Ferguson] approached the problem of parsimony by considering a single variable, represented by a point, and asking himself what was its most parsimonious description. He suggested that, intuitively, the most parsimonious description results when one of the*

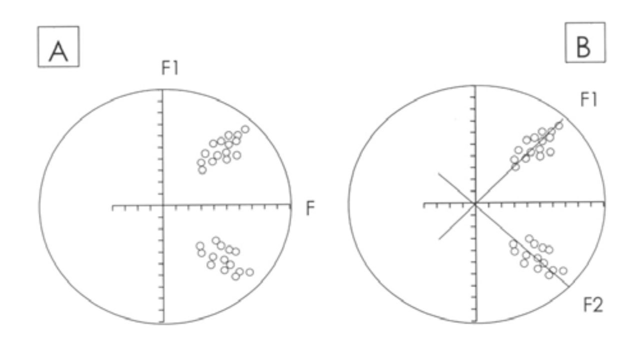


Abbildung 1.02

Zwei Punktwolken im initialen Faktorensystem F1, F2 (A)
und nach einer orthogonalen Rotation zur Einfachstruktur (B)

Bei einem z. B. zweidimensionalen Koordinatensystem mit zwei Punktmengen im Abstand eines rechten Winkels sind nach Annäherung der Koordinaten an deren Schwerpunkte die einzelnen Variablen nur durch je einen Faktor mathematisch beschreibbar, während sie für die Beschreibung der Variablen nach der initialen Lösung zwei Faktoren zu berücksichtigen waren. Mit anderen Worten: Ergeben sich zwei Faktoren, dann wird diejenige Lösung für einfacher gehalten, bei der die Variablen mehrheitlich auf nur einem der beiden Faktoren laden.

Den Vereinfachungsschritt, der bei der Zuteilung von Faktoren zu einzelnen Variablen aus ‚Sparsamkeit' vorzunehmen sei, hält man im Übrigen für eine zwangsläufige Fortsetzung der Vereinfachungsstrategie, die zur Reduktion der Variablenvielfalt auf wenige Faktoren führte: *„Im Faktorenproblem bestand das Streben nach Einfachheit darin, eine möglichst kleine Anzahl von Faktoren zu finden … Beim Rotationsproblem handelt es sich darum, innerhalb des… festgelegten… Faktorenraums das Zueinander der Variablen und Faktoren so einfach als möglich zu gestalten"* (Überla, 1971,p. 176). Am Prinzip der Vereinfachung im Sinne von ‚weniger ist besser als mehr', dem die Faktorenextraktion folgte und bei ihr mit Recht, wird also bei der anschließenden Faktorenrotation festgehalten, wo sie nicht rechtens ist, so als ob man ein Prinzip, das ein Problem lösen hilft, auf alle anderen Problemlösungen blindlings zu übertragen hätte.

---

axes passes through the point. It is seen that when the reference frame is rotated so that one of the axes approaches the point, the product of the two co-ordinates grows smaller…" (p. 169).

03. Folgen des Irrtums

Indem bei der Datenverarbeitung eine Einfachheit durchgesetzt wird, die dem formalen Modelldenken entgegenkommt, handelt man sich für das Verständnis der durch das Modell abzubildenden Sachverhalte schädliche Folgen ein. Überla gibt dafür ungewollt ein lehrreiches Beispiel: Bei 90 Männern wurde im Laufe einer halben Stunde sechs Mal der systolische und diastolische Blutdruck gemessen. Die initialen Ergebnisse einer Faktorenanalyse der 2 mal 6 = 12 Variablen sind in Abbildung 1.03A dargestellt, Abbildung 1.03B enthält die Ergebnisse einer Varimax-Lösung dieser Faktoren, die SS orthogonal optimiert. Die Lösung 1.03B stellt Überla als die einzig sinnvolle hin, da, wie er meint, die Variation der Messungen *„im wesentlichen von zwei Einflussgrößen bestimmt wird, eben dem systolischen und dem diastolischen Blutdruck"* (p. 265).

Tatsächlich sind ‚systolisch' und ‚diastolisch' keine ‚Einflussgrößen', sondern lediglich verschiedene Momente der Blutdruck-Verhältnisse innerhalb des pulsierenden Blutumlaufs. Blutdruck ist eine Einheit mit lokaler und zeitlicher Varianz. Im Moment der Systole und Diastole schlägt sich Blutdruck auf jeweils verschiedenem Niveau nieder analog dem Luftdruck in verschiedenen Höhenlagen. Eine therapeutische Regulation durch Medikamente ist nicht auf entweder systolische oder diastolische ‚Einflussgrößen' ausgerichtet, sondern auf eine Verbesserung des Blutdrucks als Systemmerkmal (s. Näheres im *Journal of Human Hypertension* und *Hypotension*). [14] Eine Faktorenanalyse der 12 Blutdruck-Variablen würde den phy-

[14] Beim Einsatz der exploratorischen Faktorenanalyse durch Internisten zur Ermittlung des Insulin Resistance Syndrom (ISR), zu dem der Blutdruck gehört, wurde zum Leidwesen der Forscher festgestellt, dass der systolische und diastolische Blutdruck nicht auf dem gleichen Faktor luden, obgleich „…*systolic and diastolic blood pressure are more strongly associated with each other than they are with other components of the insulin resistance syndrom, something which most clinicians would expect"* (Lawler et al. (2004, p. 6). Man half sich dadurch, dass man nur eine der beiden Blutdruckvariablen im ISR-Variablensatz beließ, um dem Methodenartefakt auszuweichen: *„Therefore, the evidence for inclusion or exclusion of hypertension in the definition of the syndrome is based on whether one or two blood pressure measurements are included in the model rather than on any sound clinical or pathophysiological reasoning* (Lawler et al., 2004, p. 1016).

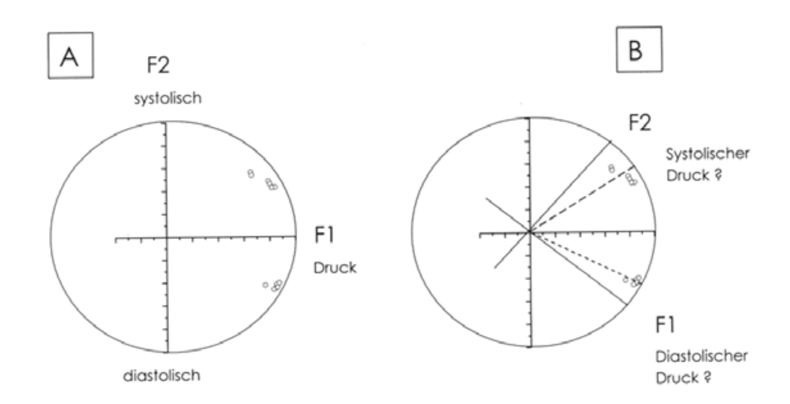


Abbildung 1.03

Initiale Faktorenlösung der Blutdruck-Daten (Überla) (A)
und Rotation zur Einfachstruktur (B) ---- orthogonal - - - schiefwinklig

siologischen Sachverhalt dann am besten abbilden, wenn die miteinander hoch korrelierende Varianz ‚systolisch-diastolisch', d. h. die Einheitlichkeit des Geschehens, durch einen Generalfaktor zur Darstellung kommt. Der zu fordernde Generalfaktor ist, wie Abb. 1.03A zeigt, in der Initial-Lösung durchaus vorhanden und geht durch die SSM-Rotation (1.03B) verloren.

Eine Differenzierung nach ‚systolisch' und ‚diastolisch' legt indessen der zweite initiale Faktor nahe. Er ist wie der Generalfaktor aus physiologischen Gründen zu erwarten, denn die beiden Blutdruckmomente zeigen zusätzliche Varianz, die auf besondere Bedingungen zurückzuführen ist. So reagieren systolische Werte auf äußere Einflüsse generell etwas stärker als diastolische. Die systolische Varianz ist bei circadianen Messwertreihen größer als die diastolische (Halberg, 1980, Fig. 8, p. 552), beim Altershochdruck steigen die systolischen Werte stärker an als die diastolischen usw. Offenbar wird die Eigenvarianz der beiden Messmomente in der initialen Lösung durch den zweiten Faktor repräsentiert, dem ein geringerer Anteil zukommt als dem ersten Faktor.

Analoges findet man auf dem Gebiet der faktoriellen Intelligenzforschung. Auch für die Intelligenz wird theoretisch ein Generalfaktor *g* gefordert, der in aller Regel in der initialen Lösung mit großer Annäherung bereits vorhanden ist. Durch eine SSM-Rotation geht *g* verloren. Da aber ein *g*-Faktor theoretisch zu erwarten bleibt,

wird dieser – meist in umständlicher Weise – wieder hergestellt [15], etwa durch Orthogonalisierung der zunächst schiefwinklig rotierten Primärfaktoren nach einem berühmt gewordenen Verfahren von Schmid-Leiman (1957). So gelangt man zu Faktoren ‚zweiter Ordnung' und damit zu einer ‚Hierarchisierung' der Modellkomponenten. Damit aber werden theoretische Folgeprobleme ignoriert bzw. in Kauf genommen. [16]

Prägnant beschreibt A. Jensen die Wirkungen einer SSM-Rotation auf den initialen ersten Faktor von Intelligenztest-Analysen: *„…The tests are all positively correlated and therefore all had some factor in common – a general factor, or g. The general factor that was so prominent in the analysis depicted in Figure 3.3 [showing the initial solution] seems to have disappeared from Figure 3.4 [showing the simple structure-rotated solution] as a result of rotating the factor axes. Actually, it has simply been dispersed (or redistributed) among the rotated factors… So if you ask where g went , the answer is that it has been divided up and lies 'hidden' among all of the tests' smaller loadings on all of the orthogonally rotated factors. Its variance has not disappeared, it has simply been obscured by being dispersed throughout the whole factor matrix."* (Jensen, 1998, p. 66). Pett et al. (2003, p. 143) zitieren Autoren, die das Problem gesehen haben: *„Nunnally and Bernstein (1994) … warn against prematurely concluding that, based on a Varimax-solution, a general factor is absent because Varimax is designed to eliminate general factors (Gorsuch, 1983). Comrey and Lee (1992) suggest that the researcher avoid including too many factors in a Varimax rotation solution because it tends to overinflate the importance of lesser factors. Although the authors do not indicate how many factors are 'too many' they point out that trial and error is the only way to arrive at the appropriate number of*

[15] Schon 1958 wurde dieses Verfahren von britischen Autoren karikiert, die sich dem Thurstone-Dogma noch nicht verpflichtet *fühlten: "It has been said that what was thus thrown out of the door [the general factor] returned through the window: for correlated factors in turn give rise to a second order factor, and this is virtually the general factor of the centroid … under another name"* (Hamilton, 1958), p.167).

[16] Dass mit hierarchischen Modellen schwerwiegende Folgeprobleme entstehen, kann hier nur angedeutet werden. ‚Hierarchie' suggeriert eine logische Gliederung nach niedrigeren und höheren Ebenen. Die bio-psychologische Realität aber operiert nicht nach Regeln der Begriffseinteilung. Was diese Realität tatsächlich aufweist, sind umfassendere und weniger umfassende Einflussgrößen, die *simultan* zusammenwirken. Die umfassendere Einflussgröße, z. B. die allgemeine Intelligenz *„g"*, ist auch ohne Mitwirkung der weniger einflussreichen Bedingungen funktional manifest (ähnlich eine kritische Bemerkung von Revenstorf, 1976, p. 313). Es gibt auf der funktionalen Ebene nichts begrifflich Über-und Untergeordnetes, sondern nur Umfassenderes (Durchschlagenderes) und Spezielleres (Modifizierendes).

factors.“ Dass man den erkannten Fehler durch ein „Versuch-und-Irrtum“- Verfahren zu korrigieren empfiehlt, kennzeichnet die missliche Lage.

04. Fehleranalysen im Detail

- Zur Abwertung initialer Faktorstrukturen

Thurstone und alle, die seiner ‚amerikanischen Schule‘ folgten, hielten das initiale orthogonale Faktorenmuster, das eine Zentroid- , Hauptachsen- oder Hauptkomponentenanalyse generiert, für „im allgemeinen nicht interpretierbar“ (Überla, 1971, 175), zudem für instabil. Zur vermeintlichen Instabilität der initialen Faktoren äußert sich Überla: *„Sie [die initialen Faktoren,] … ändern sich von Stichprobe zu Stichprobe … da neue Variablen die Gewichte oft entscheidend verschieben und die Position der Achsen dadurch abändern“* (p. 175). Dass faktorielle Invarianz bei initialen Faktoren ein Problem sei, dem man nur durch Transformationen beikommen könne, ist indessen eine unbegründete Behauptung. In der Literatur zur Faktorenanalyse konnte ich keinen Befund ausmachen, wonach sich rotierte Faktoren als invarianter erwiesen hätten als unrotierte - ein Defizit, das auch Andresen (1998) konstatiert. [17] Stattdessen finden sich durchaus Klagen über einen Mangel an Invarianz bei *rotierten* Faktoren (Fittkau, 1968, Butler, 1969).[18]

Widerspruch ist auch gegenüber der Behauptung anzumelden, initiale Faktoren seien nicht interpretierbar. Da die Variablen zumeist auf mehreren initialen Faktoren laden, so heißt es, könne man nicht wissen, was eine Variable, die mehrere Faktoren aufweist, inhaltlich darstelle. Doch darf man von zwei Interpretationen eines Sachverhalts nicht unbesehen diejenige für richtig halten, die auf Anhieb verständlicher erscheint. Man müsste sonst am geozentrischen Weltbild festhalten und das kopernikanische verwerfen.

Dass initiale Faktoren SS-transformiert werden sollen, wird auch damit begründet, dass der herausragende ‚Generalfaktor‘ ein methodisches Artefakt sein müsse, da

---

[17] *„Für die Aufgaben der Konstruktexplorierung im Bereich der Fragebogenskalen zur Persönlichkeit steht der entscheidende Beweis der Überlegenheit der Einfachstruktur-optimierenden Methode im Hinblick auf definitiv verbesserte variablenstichproben-übergreifende Strukturreplikationen noch aus.“* (Andresen, 1998, p. 74).

[18] Fittkau, 1968, p. 110: *„ …die Ergebnisse jedes der analytischen Rotationsverfahren [sind] …nicht invariant gegen Hinzunahme weiterer Variablen oder Ersetzen einiger Variablen durch andere.“* Butler, 1969, p. 13: *„ the simple structure concept does not solve one of the most crucial and fundamental problems of factor analysis, the problem of the likelihood of factorial invariance.“*

ein Generalfaktor bei sehr verschiedenen Datensätzen extrahiert werde. Überla gibt eine verbreitete Auffassung wieder, wenn er sagt: *„Die [initialen] Faktoren …weisen eine willkürliche Varianzaufteilung auf. Die Varianzaufteilung ist nicht aus den Daten abgeleitet, sondern der Methode inhärent"*. Hier wird einer rechnerisch ermittelten Regelmäßigkeit ‚Willkür' oder Zufälligkeit zugetraut, anstatt zu vermuten, dass eine Faktorenanalyse unter den Varianzquellen, die den untersuchten Phänomenen zugrunde liegen, sehr oft tatsächlich eine dominierende mit größtem Aufklärungsanteil hervorbringt, während Varianzquellen ab Faktor 2 in der Regel bedeutend weniger Varianz aufklären. [19]

Aufschlussreich ist in diesem Zusammenhang, dass inhaltsorientierte Forscher trotz solcher Warnungen der Lehrbuchautoren den ersten unrotierten Faktor einer Faktorenanalyse oft durchaus als Endergebnis akzeptieren. Manch ein Intelligenzforscher entdeckt im unbehandelten ersten Faktor den Niederschlag eines erwarteten Generalfaktors der Intelligenz (expliziert bei Jensen, 1998, p. 65-68). Maxwell (1972) beschränkte sich bei einer Varimaxrotation von Intelligenzfaktoren auf die Faktoren F_2 bis F_5 und ließ den initialen Faktor F_1 unrotiert. Auch erkennen Persönlichkeitsforscher bei der Analyse von Fragebogen-Daten oft schon im ersten Faktor das Konstrukt, das sie faktoriell operationalisiert sehen wollen (angefangen mit Hamilton, 1960, Lumsden, 1961, Leventhal & Stedman, 1970). Durch eine SSM-Rotation geht in der Regel die erwünschte initiale Lösung verloren und damit die Repräsentanz des Konstrukts, das operationalisiert werden soll. Um den ersten initialen Faktor davor zu bewahren, mit einer SSM-Rotation zu verschwinden, greifen manche Faktoristen zu einem Trick: Sie schließen bei einer Wiederholung der Analyse Variablen mit unliebsam hohen Ladungen bei einem zweiten und dritten Faktor aus, womit der zweite oder dritte usw. Eigenwert unter das Ausschlusskriterium von Kaiser und Guttman (= 1.0) gedrückt wird. So entgeht man ohne Verletzung der Konvention dem vermeintlichen Zwang zur Faktorenrotation und seinen konstrukt-destruierenden Folgen - ein formalistisches Vorgehen, das eher der Ausnutzung juristischer Gesetzeslücken ähnelt als dass es der Ermittlung sachlicher Gegebenheiten dient. Der Trick liegt als solcher offen zutage, aber niemand nimmt daran Anstoß. [20]

---

[19] Die Behauptung, dass die Lage der Achsen einer Initiallösung ‚willkürlich' sei und deshalb ohne interpretativen Wert, stammt von Thurstone (1934): *"A characteristic of the multiple factor problem is that the location of the axes is arbitrary and that hence the factorial components are to that extent arbitrary and without fundamental psychological significance."*

[20] Ein typisches Beispiel für das beliebte Ausschließen von Items, mit dem man dem vermeintlichen Zwang zur SS-Rotation entgehen möchte, gibt Thalbourne, 1998. Groner & Groner (1991) beschränken sich sogar nur auf den ersten von 17 (!) initialen Faktoren mit Eigenwerten größer als 1 allein aus

Methodisch sauberer gingen Gibbons & Hederer (1992) vor, die das Problem der oft sich aufdrängenden Deutbarkeit des initial auftretenden Generalfaktors mit einem sogenannten Zweifaktorenmodell („bi-factor model“) zu lösen suchten. Beim Zweifaktorenmodell wird ein Generalfaktor (F_1) vorausgesetzt bzw. akzeptiert, an dem alle Variablen partizipieren, während bei jeder einzelnen Variablen einer von mehreren weiteren Faktoren ($F_2 \dots F_s$) zugelassen wird. Es zeigte sich, dass durch das Rotationsmodell, das den Generalfaktor konserviert, bei der Verarbeitung von Daten z. B. aus einem Wissenstests und aus einem Depressionsfragebogen ein signifikant besseres Passungskriterium erreicht wurde als durch SSM-Rotationen. [21]

Um den Generalfaktor zu retten, wenn ein solcher gebraucht wird, aber durch Rotation verschwindet, sind schon sehr früh besondere Rotationstechniken entwickelt worden. Die orthogonale Quartimax-Rotation von Neuhaus & Wrigley

---

Interpretationsgründen. Sabouria et al. (1990) informieren über eine jahrzehntelange Kontroverse, die um die Frage des Verlustes an erwünschter und behaupteter Eindimensionalität anhand eines unrotierten ersten Faktors entbrannte (im Zusammenhang mit der Entwicklung der Dyadic Adjustmment Scale (DAS) von Spanier). Bei Gangestad & Snyder (1985) und Snyder & Gangestad (1985, 1986) findet sich eine aufschlussreiche Begründung zur Rettung des Konstrukts ‚Self-monitoring‘, das durch Simple Structure-Rotationen verloren gehen würde. Dilemmata dieser Art werden zunehmend durch methodische Kompromisse zu entschärfen versucht, zu denen das Extrahieren von Faktoren zweiter Ordnung (führt zu einem ‚hierarchischen‘ Modell) und die Überprüfung diverser Modellmöglichkeiten durch konfirmatorische und Strukturgleichungsverfahren gehören (Undheim & Gustavsson, 1987: „Restoring general intelligence“). Die eigentlichen Ursachen der Probleme werden dadurch verschleiert.

[21] Hamilton (1958) bediente sich der seinerzeit noch aktuellen Methode der ‚simple summation‘, die ihm aus einem Angst-Fragebogen einen Generalfaktor der Angststärke lieferte (F_1) neben einem schwächeren Faktor F_2 der Angstebene (psychische vs. somatische Symptom-Dominanz). Ein Beruhigungspharmakon, Benactyzin, das den am Versuch teilnehmenden Patienten gegeben wurde, führte, wie erwartet, zu Veränderungen bei F_1, nicht bei F_2. Der Verfasser kritisiert Thurstones SS-Verfahren, bei dessen Verwendung faktorielle Deutungsprobleme auftraten und bei dem der Behandlungseffekt faktoriell nicht mehr als ein eindeutig angstreduzierender in Erscheinung trat. Eine Varimin-Reanalyse der mitgeteilten Korrelationsmatrix konnte die Kritik und Faktorendeutung Hamiltons bestätigen (unveröffentlichte Reanalyse des Verfassers).

(1954) kann dies in gewissem Maße leisten, wenn man die offensichtlichen Schwächen, die diese Methode mit sich bringt (Gorsuch, 1974, p. 191), mit in Kauf nimmt einschließlich der weiteren Simple-Structure-Folgen, die Quartimax nicht beseitigt.

Dennoch halten es Überla und alle, die seine Lehrbuch-Ansichten teilen, für unausweichlich, die vermeintlich ‚unverständlichen' und ‚willkürlichen' initialen Faktorenlösungen in vermeintlich verständlichere und stabilere Lösungen zu überführen. SS wird als die einzig vertretbare oder sogar einzig mögliche Leitlinie hingestellt, nach der sich das Koordinatensystem auszurichten habe. Alternativen werden nicht diskutiert, etwas anderes als das SS-Prinzip hält man offenbar für nicht denkbar.

In der Frühzeit der Faktorenanalyse fand die amerikanische Schule Thurstones noch Konkurrenz und Kritik durch die „britische Schule" unter der Führung von Cyril Burt, der in den Initiallösungen der Faktorenanalyse, in den „bipolaren Strukturen", wie er sie nannte, das Vorhandensein wertvoller Informationen erkannte: *„…to the experienced factorist both the regularities and irregularities [of the pattern of signs] will yield considerable insight into the data he is analysing, even without any further rotation or analysis"* (Burt, 1954, p. 16). Burt nahm routinemäßig zwar auch eine Umwandlung der bipolaren Initialfaktoren zugunsten einer unipolaren Lösung vor, die er ‚Gruppenfaktorlösung' nannte. Doch damit entfernte er sich von der Initiallösung weniger weit als die amerikanischen SS-Faktoristen. Zurückhaltung gegenüber einer verfälschenden Faktorenrotation findet man auch u. a. bei Burroughs & Miller (1961, p.35-37): *„…the subsequent rotations are apt to obscure [the objective and dichotomous classification based directly on the data]"* (p. 36). Ein längeres Andauern der Konkurrenz zwischen den britischen und amerikanischen Faktoristen, zu der es nicht kam, hätte die Fehlentwicklung, die die Thurstonesche Position nach sich zog, vermutlich verzögert oder sogar verhindert.

- Fehldeutung der ‚positive manifold'

Intelligenztests korrelieren untereinander generell positiv (‚positive manifold'). Die initialen Faktorlösungen von Test-Interkorrelationen weisen jedoch - wie bei den Blutdruckdaten - außer einem unipolar positiven Generalfaktor F_1 (= *g*) in der Regel nur noch bipolare Faktoren F_2, F_3 usw. auf, bei diesen also auch negative Ladungen. Die Intelligenzforscher der ersten Stunde, die als Faktoren ‚Fähigkeiten' erwarteten, hielten negative Ladungen für widersprüchlich. Sie waren davon überzeugt, dass Faktoren aus Intelligenztest-Untersuchungen nur positive Ladungen, allenfalls Null-Ladungen (= keine Fähigkeit), aber keine negativen Ladungen haben könnten. Negative Ladungen hätte man als Leistungsminderung zu interpretieren, was dem Begriff der Fähigkeit widerspreche. Eine Umwandlung der

bipolaren zur nur-positiven Ladungsverteilung erschien also notwendig. Da man glaubte, die positiven Interkorrelationen durch einheitlich positive Ladungsvorzeichen auch in der Faktorenmatrix darstellen zu müssen, hielt man die Transformation zur SS, die dies bewirkte, für zusätzlich gerechtfertigt (Thurstone, 1947, p. 341) [22] .

Anhand der Blutdruck-Untersuchung wurde deutlich, dass die aufzuklärenden Sachverhalte durch SSM verfälscht werden. In ähnlicher Weise wird das faktorielle Endergebnis verfälscht, wenn man die positive Mannigfaltigkeit der Intelligenztest-Korrelationen mithilfe der Faktorenmatrix zu konservieren sucht. Hier führt nicht mehr das mathematische Modell zu fehlerhaften Sachkonzepten, vielmehr nehmen umgekehrt vorgefasste Sachkonzepte (der Fähigkeitsbegriff) Einfluss auf die mathematische Modellierung (hier: die Vorzeichenmanipulation).[23] Überla war bei seiner Blutdruck-FA offenbar so wie Forscher bei einer Intelligenz-FA davon ausgegangen, dass man nur positive Ladungen zulassen dürfe, da der niedrigste Wert einer Hg-Skala Null und nicht negativ ist. Fälschlicherweise überträgt er dabei die Eigenschaften der Metrik, die den manifesten Beobachtungen angemessen ist - beim Blutdruck wie bei Intelligenzleistungen ist eine Verhältnisskala angemessen -, auf die Metrik der latenten Bedingungen, auf der nur die Eigenschaften einer Intervall-Skala vorliegen. *„…rigorous measures, such as direct counts, latency, or duration, are excellent measures if used as descriptions of behavior but may become arbitrary*

---

[22] *„It is …natural to postulate that when a unique simple structure is found for a battery of tests of mental abilities, then the non-vanishing entries in the factorial matrix are positive*" (Thurstone, 1947, p. 341). Thurstone meint an dieser Stelle, dass die einschränkende Bedingung der „positive manifold" zur Anwendung des Prinzips der „simple structure" nicht notwendig sei. Doch sei die Rotation zur Einfachstruktur vor allem dann angezeigt, wenn *„the factor loadings shall be positive or zero". (23)* Berneyer (1957) referiert in diesem Sinne: "*The different methods of [factor] analysis [of mental aptitudes] yield factors which have negative loadings … Such factors, so Thurstone contends, must be devoid of 'scientific meaning'. They do not permit us to 'interpret the various tests as functions of the mental aptitudes which those tests elicit'*" (p. 23). Auch C. Burt teilt uneingeschränkt Thurstones „positive manifold"-Idee: *"…an ability for x is by definition a dispositional property that facilitates doing x, i.e., it denotes a positive and never a negative tendency. Hence we shall be compelled to seek factors with positive saturations only" (*Burt, 1954, p. 18).

[23] Die verfälschende Wirkung einer Simple Structure-Rotation bemerkten auch Highmore & Taylor (1954) bei der Analyse von Sporttest-Daten: *„…the ‚basic' factor, representing general athletic ability (in which we are primarily interested), necessarily disappears, and the group factors [of simple structure rotation] show little relation to the classification indicated by the [initial] bipolar matrix"* (p. 4).

metrics if they are used to infer some psychological construct“ (Kazdin, 2006). Nach der mit fünf Beiträgen geleisteten Diskussion zur ‚arbitrary metrics‘ (American Psychologist, 61, 2006) dürfte heute Einigkeit darüber zu erzielen sein, dass auch die Metrik, die sich auf der latenten Ebene von Faktorladungen ergibt, ‚arbitrariness‘ von Nullpunkten und Vorzeichen nicht nur erlaubt, sondern erfordert. Schon Vukovich (1967) vertrat in diesem Punkte Liberalität: „*Mit den Messwerten ist jeder Umgang erlaubt, der ihren Abbildungscharakter nicht beeinträchtigt. Wenn die empirischen Vergleichsoperationen nichts über die Nullpunktslage und die Größe der Einheiten eines Maßstabes aussagen, kann man sie frei wählen und wird sie bequemerweise so festsetzen, dass größere Bedingungszusammenhänge übersichtlich zu beschreiben sind.*“ (p. 114).

Kurz: Ein initialer Generalfaktor darf nicht als Methodenartefakt (Überla) abgewertet und beseitigt werden. Dass in einer Phänomendomäne unter jeweils differenzierbaren Varianzquellen eine mit Haupteffekten hervortritt, kann der Normalfall sein. Zur Erklärung der verbleibenden Varianz, welche weitere Faktoren aus der Residualmatrix der Korrelationen extrahierbar macht, wird man in der Regel zusätzliche Bedingungen annehmen dürfen, die den Haupteffekt orthogonal modifizieren. Der Generalfaktor dient den nachfolgend extrahierten Varianzquellen als Referenzquelle. Er entspricht dem Mittelwert = 0 einer standardisierten Werteverteilung, von dem ausgehend sich positive und negative Abstände ergeben. [24] Das Vorkommen von negativen neben positiven Ladungen bei einem zweiten, dritten usw. Faktor bedeutet also nur, dass der Einfluss zusätzlicher Varianzquellen in ihrer Größe zu den Effekten der Hauptvarianzquelle in Beziehung gesetzt werden. Aufgrund der Intervalleigenschaft der zusätzlichen Varianzquellen ergibt sich dann eine bipolare Verteilung. Ähnliche Überlegungen wurden bereits von Thompson (1963) angestellt, sie fanden kaum Resonanz. [25]

Die vorliegende Kritik an der Einschätzung der „positive manifold“ durch Thurstone und Nachfolger sollte nicht als Empfehlung missverstanden werden, die Initiallösung von Faktorenextraktionen als Endlösung einer Faktorenanalyse einzuschätzen. Auch Initiallösungen erfordern meistens Verbesserungen durch

---

[24] Dies wurde geprüft und bestätigt gefunden mit Daten aus dem sportlichen Leistungsbereich (Zehnkampfleistungen, s. Kapitel 4).

[25] *“If they are able to choose one or the other, …psychologists tend to prefer uni-directional to bi-polar measurement, probably as a result of the prestige of such measures as the standard metre in physics. … Thurstone expressed a preference for a positive manifold (without, in the writer’s opinion, giving a fully convincing explanation)…Bi-polar and unidirectional measurement are both needed in psychology”* (Thompson, 1963, p.22). Die letztere Behauptung findet sich ausführlich in Thompson, 1962, begründet.

eine Faktorenrotation. Dies wird im folgenden Abschnitt 05. und unter 08. weiter ausgeführt.

05. Neuorientierung

- Basisbedingung natürlichen Geschehens: Komplexität

Am Beispiel der Blutdruck-FA wurde deutlich, dass die Gesamtvarianz der erhobenen Variablen auf zwei latente Quellen aufzuteilen ist, die eine repräsentiert den Blutdruck unabhängig von Varianzen der Messzeitpunkte, die andere bringt den Varianzanteil hervor, den die Messzeitpunkte zur Gesamtvarianz beitragen. Damit werden die empirischen Gründe der vorliegenden Kritik am SSM transparent. Man kann sie für nahezu trivial halten: Wahrscheinlich auf allen Gebieten empirischer Forschung liegt den jeweils beobachteten Variablen eine Mehrzahl zusammen wirkender Bedingungen zugrunde. Die SSM-orientierte Faktorenanalyse operiert an dieser Trivialität vorbei. Unbestritten bleibt zwar der Sparsamkeitserfolg des ersten Schritts einer Faktorenanalyse, da den manifesten, untereinander korrelierten Variablen eine geringere Zahl latenter Varianzquellen (faktorielle ‚Bausteine') zugeschrieben wird. Doch wird dieser Erfolg verspielt, wenn man den manifesten Variablen durch Faktortransformation möglichst jeweils nur einen latenten Baustein oder ein Minimum von diesen zuweist. Das Gegenteil wäre zu fordern (mehr Argumente hierzu in Abschnitt 08).

Symptomatisch für den Irrtum, auf dem das SS-Ideal beruht, ist die Beobachtung, dass in der bisherigen faktorenanalytischen Praxis Einfachstrukturen so gut wie nie gefunden wurden. Egal welche SS-Rotation dafür verwendet wurde. Alles bisherige Ringen um Simple Structure war eine Art Don Quichoterie. Schon der Klassiker Harry Harman musste zugeben: *"An orthogonal uni-factor solution is practically impossible with empirical data, and not very likely even when the factors are permitted to be oblique. Nonetheless, … it is towards that end that the simple structure principles are proposed for the multiple factor solution"* (Harman, 1968, p. 99). [26]

---

[26] Kritik am Prinzip der Einfachstruktur ist nur in Ausnahmefällen laut geworden: *„Die Entscheidung zwischen den vielen mathematisch möglichen Lösungen wurde durch eine formale Regel, das Minimalprinzip („simple structure") getroffen, dem jede theoretische Begründung fehlt. Frei von jedem Zwange, grundlegendere Überlegungen anstellen zu müssen, wurde nun Massenarbeit geliefert…Der Erfolg dieses Vorgehens war, wie vorauszusehen, eine Überproduktion von oberflächlich definierten Faktoren"* (Meili, 1969, p. 278). Auch Revenstorf (1976) hielt Einfachstrukturen, zumal bei „großen Merkmalzusammenstellungen (Fragebögen)… für im allgemeinen unwahrscheinlich", weil die *„Merkmale mehr und mehr alle möglichen Kombinationen von*

Das Ziel der Faktorenanalyse ist dem der multivariaten Varianzanalyse (MANOVA) ähnlich. Der Hauptunterschied besteht darin, dass der Faktorenanalytiker mit *latenten unabhängigen* Variablen (UVn, Varianzquellen, ‚Faktoren') zu tun hat, die den manifesten, beobachteten abhängigen Variablen (AVn, z. B. Testvariablen) zugrunde liegen sollen. Latente Variablen sind zu Beginn einer faktorenanalytischen Untersuchung unbekannt oder ungesichert und sollen aufgedeckt oder gesichert werden. Ein Varianzanalytiker dagegen hat nicht nur mit manifesten abhängigen, sondern auch mit manifesten *unabhängigen* Variablen zu tun, die ihm bekannt sind und die er experimentell manipuliert. Sein Analyseverfahren dringt indessen nicht bis zur Ebene der *latent* wirksamen unabhängigen Variablen der ‚Black Box' vor, d. h. nicht bis zu den Prozessen, die durch die Manipulation der manifesten UVn „submanifest" aktiviert werden und als *proximale* Determinanten die abhängigen Variablen hervorrufen oder verändern. Das ist der große Nachteil varianzanalytischen Forschens.

Ein weiterer Unterschied zwischen der konventionellen Faktorenanalyse und der MANOVA ist erwähnenswert: Die MANOVA berücksichtigt auch ‚strukturelle' Bedingungen der untersuchten AVn, also Komplexität, die dort als *Interaktion* behandelt wird. Allerdings handelt es sich dabei nur um eine Interaktion manifester Variablen. Die SSM-orientierte Faktorenanalyse dagegen ignoriert Interaktionen auf der latenten Ebene. Faktorielle Einheiten, die SSM-konzipiert werden, bleiben untereinander ohne Beziehung.

Mit einem vorurteilslosen Rückgriff auf unser Elementarwissen hätte das SS-Prinzip von vorne herein suspekt erscheinen können, denn den operationalisierten Variablen wissenschaftlicher Forschung liegt auf einer tieferen Ebene generell ein Zusammenwirken von Komponenten zugrunde. Atome bestehen aus Protonen, Elektronen und Neutronen, Salz aus Natrium und Chlorid, Gene haben nur im-

---

Faktorausprägungen aufweisen können. In diesem Falle streuen die Merkmale der Variablenkonfigurationen über den ganzen Faktorenraum und eine Einfachstruktur im Thurstone'schen … Sinne ist nicht mehr festzustellen" (p. 321). Unzählige Beispiele aus dem Alltag der Faktoreninterpretation ließen sich anführen: „*Es ist klar, dass die Einfachstruktur zu wünschen übrig lässt*" (Bierhoff, 2000); „*… ist trotz der vergleichsweise hohen Interkorrelationen nicht von einer Einfachstruktur auszugehen*" (Schaper & Baumgart, 2002); „*Insgesamt erscheint die Prägnanz der Einfachstruktur eingeschränkt"* (Beauducel et al., 2003), „*…keine Einfachstruktur nach dem Bargman-Test*" (Herzberg, P. Y., 2002); *„Eine akzeptable Einfachstruktur der Faktorladungsmatrix ließ sich weder durch orthogonale, noch durch schiefwinklige Rotationen der drei Hauptachsen erreichen"* (Schmitt, M. , 2000); *„Die Lösung lässt nicht auf eine Einfachstruktur schließen"* (Lambert et al.. 2002).

mer in komplexer Vielfalt eine Auswirkung (‚Polygenie'). [27] Die Wahrnehmung einer einzelnen Farbe ergibt sich aus der Kombination von elektrischen Erregungen, die von drei Zapfenpigmenten produziert werden. Sätze bestehen aus Wörtern, Wörter aus Morphemen, Morpheme aus Phonemen. Jedem Phonem liegt ein Komplex phonetischer Merkmale zugrunde. Auf die Faktorenanalyse angewandt: Man sollte erwarten, dass den manifesten Variablen einer Untersuchung in aller Regel eine Mehrzahl zusammen wirkender latenter Varianzquellen zugrunde liegt.[28]

Komplexität, das Zusammenwirken bedingender Komponenten, verträgt sich durchaus mit Occams Ideal der Einfachheit, ja sie geht aus diesem als Konsequenz unmittelbar hervor. Wenn verfügbare Einheiten in einem System kombiniert werden können und dadurch neue adaptive Leistungen entstehen, die vom Organismus gebraucht werden, erübrigt sich die aufwendige Herstellung weiterer spezieller Einheiten für neue Leistungen. Die Physiologie der Farbwahrnehmung kommt mit drei sensorischen Empfangsrezeptoren (Pigmente) aus, für die Farbqualitäten lila, ocker, ultramarin usw. waren keine besonderen Rezeptoren zu entwickeln. Die jüngsten Untersuchungen zur Evolution der Sprache kulminieren in der Erkenntnis, dass selbst die höchsten mentalen Leistungen des Menschen durch eine erfolgreiche Kombinatorik verfügbarer prozessualer Ressourcen zustande kommen. Für neue Funktionen wie die der Sprache bei Hominiden ein von Grund auf neues Entwicklungsprogramm einzuleiten, wäre nach Bates et al. (1992) [29] und Gould et al. (2002) für die Evolution eine Verschwendung gewesen. Wolf Singer (2003) wird nicht müde, die menschliche Großhirnrinde mit der Fä-

---

[27] Einfachste anthropometrische Merkmale wie Körpergröße und Haarfarbe beruhen auf einem Zusammenwirken mehrerer Gene.

[28] Methodische Ansätze, die der Kombinatorik latenter Größen bei der Produktion manifester Einheiten Rechnung tragen, findet man außerhalb faktorenanalytischen Denkens, so z. B. in Wiley's ‚Latent Partition Analysis': *"…it is assumed that the manifest categories are derived by dividing and combining the latent categories. We have found that by examining manifest categories it is possible to reconstruct information about the latent partition and about its relation to the manifest partitions."*(Wiley, 1969, p. 183).

[29] *"Language learning appears to be based on a relatively plastic mix of neural systems that also serve other functions. I believe that this conclusion renders the mysteries of language evolution … somewhat more tractable. That is, the continuities that we have observed between language and other cognitive systems make it easier to see how this capacity came about in the first place."* (Bates, E. Modularity, domain specifity, and the development of language. h*ttp://www.ecs.soton.ac.uk/~harnad/Papers/Py104/ bates-1994.html.)*

higkeit zu beschreiben, „ein kombinatorisches Spiel zu spielen“ (p. 84). Die elementaren Strukturen des Gehirns haben „gleiches Format“, sind „erstaunlich monoton aufgebaut“ (p. 44), „die Natur ist sehr konservativ… “ (p. 46). Das Hirngeschehen bringt bei fehlender Zentralisierung „seine Leistungen durch konstruktivistische Bindungen“ (= Kombinatorik) hervor (p. 75). *„[In der] funktionellen Architektur des Gehirns [ist]entscheidend, wer mit wem wie stark und ob erregend oder hemmend in Kontakt tritt*“ (p. 38). Für den Kontakt einer einzelnen Hirnzelle mit anderen ist durch 10-tausende Dendriten gesorgt. Die Übertragung von Thurstones SS-Modell auf das Hirngeschehen, die Auffassung neuronaler Einheiten als solitäre Pflöcke, würde dem Prinzip der ‚freien Bindung‘ direkt widersprechen. Das Bindungsphänomen ist für Singer das ‚Kernproblem der Neurowissenschaften‘ (p. 57). Offenbar unterliegt das Geschehen in der gesamten belebten Natur, auch sogar in der unbelebten, einer Kombinatorik von Elementen oder Bausteinen. In einer systematischen Abhandlung weist der Biologe Humberto Maturana (1998, p.158- 189) hin auf die Universalität der ‚strukturellen Determination‘ des Naturgeschehens. [30]

Das Problem der Komplexität verdient eine Diskussion auch aus einer noch umfassenderen Perspektive, entsprechende wissenschaftsphilosophische Vorarbeiten liegen vor. Malcolm Forster (1998) rollt die Problematik der Begriffe *„Parsimony and Simplicity“* auf und folgert: *„The paucity of parameters is a limited notion…There are no compelling ideas about why such properties should count in favor of one theory being closer to the truth than another“. „Whether a simple curve is preferable to some more complex alternative, or the reverse is true, has nothing to do with simplicity and everything to do with predictive accuracy“* (Forster & Sober, 1994, p. 28). Auch gehören hierher Zitate aus Nicholas Reschers *Complexity. A Philosophical Overview „…In the development of knowledge - as elsewhere in the domain of human artifice - progress is always a matter of complexification. An inherent impetus towards greater complexity pervades the entire realm of human creative effort“* (Rescher, 1998, p. 58). *„There really are no adequate grounds for supposing the ‘simplicity’ of the world’s make-up. Instead, the so-called ’principle of simplicity’ is really a principle of complexity-management“* (p. 61). Ebenso findet man bei Bechtel & Richardson (*Discovering complexity*, 1993) im Grundsätzlichen positive Resonanz: *„It is ultimately the pattern of connections in the system, and not the jobs performed by specific units in the system…, that is critical to the behavioral systems“.* Aus wissenschaftstheoretischen Gründen wird das Postulat der Einfachheit auch von Gawronski (2000) relativiert: Einfachheit ist ein „vages Kriterium“. *„Handelt es sich um Punktwolken, die durch eine mathematische Funktion beschrieben werden sollen, so mag es noch einen gewissen Konsens über die ‚einfachste‘ Kurve geben. Handelt es sich jedoch um verbal formulierte wissenschaftliche Theorien, können*

[30] Unter den zahlreichen zitierbaren Aussagen sei ausgewählt: *„Ein lebendes System ist ein strukturdeterminiertes System, und alles in diesem System geschieht aufgrund von Nachbarschaftsrelationen zwischen seinen Bestandteilen…“* (Maturana (1998, p. 184).

Urteile über die Einfachheit in Abhängigkeit von Wissensstrukturen … sehr unterschiedlich ausfallen. In diesem Sinne muss die Forderung nach Einfachheit als relativ betrachtet werden" (p. 10).

- Neues Leitmotiv zur Faktoren-Transformation: Komplexstruktur, Complex Structure Modelling , CSM).

Aus alledem folgt, dass faktorenanalytische Forschung unter ein Leitmotiv gestellt werden sollte, das das SS-Prinzip Thurstones in sein Gegenteil verkehrt. Das neue Leitmotiv ist das der Komplexstruktur. Das Ziel einer Nachbehandlung initialer Faktoren durch Transformation sollte sein, der Komplexität des latenten Geschehens, das den manifesten Variablen zugrunde liegt, uneingeschränkte Repräsentanz zu ermöglichen. Initiale Faktoren einer EFA sind so zu transformieren, dass sie das Zusammenvorkommen der gefundenen Bausteine (Faktoren) optimieren, nicht verschlechtern und verdunkeln. Dass dies mathematisch-statistisch möglich ist und sich gegenüber den konventionellen Verfahren in überraschendem Maße bewährt, wird im zweiten Kapitel dargestellt. Dort wird auch darzulegen sein, dass die initiale Faktorenlösung ein Optimum an Komplexität auch ohne Faktorenrotation schon dann hervorbringt, wenn die Analyse des Datensatzes nur zwei Varianzquellen (Faktoren) erkennen lässt (wie beim Blutdruck). Eine Komplexrotation, die bei mehr als zwei extrahierten substantiellen Faktoren erforderlich ist, sollte den Variablen Anteile *aller* faktorieller Varianzquellen zuzuordnen versuchen, die für die betreffende Domäne ermittelt werden und dies von den empirischen Gegebenheiten prüfen lassen [31]. Eine Ausrichtung auf CSM will reale Komplexität zur Abbildung führen, während die Ausrichtung auf SSM nicht-vorhandene Einfachheit zu erzwingen sucht. Das Ideal der solitären manifesten

---

[31] Die letzte umfassendere Kritik an der Faktorenanalyse im deutschsprachigen Raum (durch Holz-Ebeling, 1995) nimmt Anstoß an der Vagheit bei der Deutung von Faktoren. Die Autorin fordert mit Recht, dass die Auffassung multipler Bedingtheit („multi-UV-Bedingtheit") der untersuchten Variablen (AV) aufrechterhalten werden sollte und stellt fest, dass die Faktorenanalyse der multi-UV-Bedingtheit nicht gerecht wird. Sie schlägt deshalb methodisch ein der Varianzanalyse verwandtes Verfahren vor, das die Faktorenanalyse „ersetzen oder ergänzen" soll. Für ein varianzanalytisches Denken ist die Existenz von multi-UV, welche auf singuläre AV Einfluss nehmen, selbstverständlich. Der Vorschlag der Autorin, so unpraktisch er sonst auch ist (nach ihrem eigenen Urteil), ist interessant als Versuch einer Korrektur von Fehlern der komplexitätsfeindlichen Simple Structure-orientierten Faktorenanalyse. Die Richtung stimmt ungefähr, doch rührt die Autorin die eigentliche Ursache des Dilemmas nicht an.

Variablen (jede Variable begründe sich durch nur einen Faktor oder durch möglichst wenige Faktoren) führt zur Nicht-Einfachheit und Nicht-Sparsamkeit. CSM als Leitmodell der Faktorentransformation ist keine Absage an Newton's ‚natura est simplex' und ‚Occam's razor'. Das Rasiermesser Occams ist lediglich so anzusetzen, dass es die Strategie der Sparsamkeit, die die Natur selbst schon leistet (durch ihre ‚Bindungen', s.o.), zutage treten und nicht verschwinden lässt.[32]

Das CS-Modell will nicht als grundsätzliche Neuerung verstanden werden, denn es überträgt eine in der Wissenschaft weithin vertretene Auffassung auf ein methodisches Operationsfeld, auf dem sie bisher nicht hat Fuß fassen können. Dem multivariaten und interaktiven Geschehen wird ansonsten durchaus Rechnung getragen (zumal im Konnektionismus und in der Systemtheorie). Leider hat das abwegige Vereinfachungskonzept der SS-Analyse in einer abgeschirmten Enklave der statistischen Methodologie bis heute überdauert und die substanzwissenschaftliche Forschung fehlgeleitet. [33] [34]

---

[32] Immer wieder sind Faktorenanalytiker, die sich mit sperrig-komplexen Variablen konfrontiert sahen, an die Grenzen des SS-Prinzips gestoßen. Guilford & Zimmerman (1963) wagen eine Liberalisierung des Sparsamkeitsprinzips, um Komplexität zuzulassen: *„In general, investigators need to relax somewhat their drive to achieve parsimony. Long ago, psychology should have progressed beyond the stage in which investigators continue to look for the 'philosopher's stone'"* (p. 299). Will man Komplexität modellieren, so sind an der Sparsamkeitsforderung keinerlei Abstriche zu machen. Man hat nur anzuerkennen, dass die Sparsamkeitsstrategien der Natur andere sind als die, die Thurstone zur geometrischen Abbildung der von der Natur hervorgebrachten Variablen für geeignet hielt.

[33] Aus der Perspektive des CSM lässt sich auch die von Thurstone schon erfundene und von Cattell vehement propagierte Strategie der schiefwinkligen Rotation besser verstehen. Sie stellt ein missglücktes Spagat dar zwischen der Suche nach einer einfaktoriellen Abbildung von Variablen und der Einsicht, dass man darüber hinaus auch der Komplexität genügen muss, die im Universum der beobachtbaren Variablen natürlicherweise anzutreffen ist. *„It is unreasonable to expect that a great variety of influences operating and interacting in the same universe would be completely uncorrelated"* (Cattell & Dickman, 1962, p. 390). Der schiefwinklig rotierende Faktoranalytiker versucht deshalb, Interaktionen zwischen den Variablen durch Korrelation der Faktoren in einem nicht-orthogonalen Faktorenraum abzubilden. Die Variablen bleiben dabei Simple Structure-gebunden, und dies in besonderem Maße. Gleichzeitig wird durch die Hintertür, indem man die Korrelierbarkeit der Faktoren untereinander zulässt, einem funktionalen Zusammenhang Einlass gewährt – ein dubioser Kompromiss, den schon Guilford & Zimmermann (1963, p. 289) kritisierten: *„[This amounts to] a hollow and accidental victory for oblique methods of rotation"*), der durch Verzicht

06. Quellen des Irrtums

Was hat zur Fehlkonstruktion des *SSM* geführt? Vermutlich haben gewöhnliche Fehleranfälligkeiten des Denkens mit dazu beigetragen. Gestaltpsychologen, die sich mit den meist adaptiv nützlichen Formbildungstendenzen der Wahrnehmung beschäftigten (‚Gestaltgesetze'), warnten wiederholt auch vor ihren schädlichen Nebenwirkungen: Eine *„Behinderung des Denkens durch Anschauung"* (Witte 1974), *„Prägnanz' as an obstacle to problem solving"* (Kanizsa, 1975) scheint auch im Falle der Punktwolken im Faktorenraum vorzuliegen, die von den umliegenden Koordinaten nach den ‚Gesetzen der Nähe und Gruppierung' unwiderstehlich angezogen werden. Offenbar ist man der ‚intuitiv zwingenden' Wahrnehmungstendenz gefolgt, den Punktwolken einen Halt zu geben, was nach ‚curve fitting' -Manier mithilfe der frei drehbaren Koordinaten leicht möglich war. [35] Gorsuch (1974, p. 164 f) spricht von einer ‚visual compellingness', die Thurstone bei seinen frühen Ver-

---

auf die theoretisch unverzichtbare Unterscheidbarkeit funktionaler Einheiten erkauft wird. Die Pseudolösung des Problems mithilfe schiefwinkliger Rotation wird hier deshalb erwähnt, weil sie die Spannung verdeutlicht, die das frustrierende Streben nach Einfachstruktur zur Folge hat: Man versucht, das selbst geschaffene Übel durch fragwürdige Manipulationen zu beseitigen.

[34] Symptomatisch für das irrige SSM ist auch der Versuch von Cattell & Radcliffe (1962), die real vorhandene Komplexität, die durch eine SS-Rotation nicht beseitigt wird, durch Unterdrückung unerwünschter Varianz zu eliminieren: „*If we are right in assuming that behaviour which has a large personality factor variance will generally be factorially complex, then the unwanted common factor variance will, as a rule, be far from negligible.*" Den Autoren liegt daran, „unifactor scales" zu entwickeln und sie stellen zu diesem Zweck ein Verfahren dar, *„…which will reduce the contribution of unwanted factors by suppression"* (p. 125).

[35] Die anschauliche Prägnanz der Einfachstruktur ist zwar weniger durchschlagend als die des täglichen Sonnen-Umlaufs um die Erde, die bis zur kopernikanischen Wende gravierende kosmologische Verzerrungen zur Folge hatte. Doch scheint die Ursache der Simple Structure-Verirrung auf ähnlichen Mechanismen zu beruhen. Die komplizierten Circumplex-Nachbesserungen bei Simple Structure-rotierten Faktoren erinnern an die Zusatzberechnungen für die planetarischen Anomalien (Epizyklen), mit denen die geozentrische Vorstellung vom Kosmos konserviert wurde. Auch Quine führt als Grund für die generelle Vorliebe für Einfachheit neben dem motivationalen Wunschdenken einen basaleren Wahrnehmungsmechanismus an: "*There is a subjective selectivity that makes us tend to see the simple and miss the complex*" (Quine, 1963, p. 103).

anschaulichungen extrahierter Faktoren leiteten, welche zur Festlegung seiner oft zitierten fünf Kriterien für die SS-Rotation führten. [36]

Eine weitere Ursache für Urteilsfehler, die dem Argument des ‚leichten Interpretierens' von SS-transformierten Faktoren zugrunde liegen könnte, beschreibt Arkes (1991): *„Suppose a person adopts a quick and dirty strategy to solve a problem. Because it is so quick, it is easy to execute. This is a benefit. Because it is dirty, it results in more errors than a more meticulous strategy. This is a cost. Although the choice of this strategy results in [more errors] …, this cost may be outweighed by the time and effort saved"* (p. 487). Dies nennt Arkes, wohl nicht ganz treffend, einen strategischen Urteilsfehler (*„strategy-based judgment error"*), die Bezeichnung „cost-saving judgment error" würde besser passen.

Ähnliches findet man bei Edmonds (2002), der den universellen Hang (‚bias') zur Einfachheit einer menschlichen Schwäche zuschreibt. Denn *„…the simpler theory is not more likely to be true and is not likely to be nearer the truth…For human beings it is much easier to elaborate … [a failing] theory, or otherwise tinker with it, than to undertake a more radical shift (for example, by scrapping the theory and starting again). This elaboration may take on many forms, including … complicating the model with extra equations or rules … or using more complicated functions."* Edmonds fasst zusammen. *„…Model selection, for the sake of simplicity' is either: simply laziness … [or] due to pragmatic reasons."* Er plädiert dafür, auf „Einfachheit aus Furcht vor Neuheit" grundsätzlich zu verzichten. *„… The elaboration of [an existing] theory in order to fit a known set of data should be resisted …The lack of success of a theory should lead to a more thorough and deeper analysis than we are usually inclined to perform."*

Zur Immunität des SSM hat zweifellos auch die Macht der Worte beigetragen. Gegen einen Begriff, der sich mit der Zusammenstellung der attraktiven Attribute ‚einfach' und ‚Struktur' sprachlich Geltung verschafft, wird kaum jemand opponieren wollen. Hätte Thurstone sein Rotationsprinzip z.B. *Principle of Solitary Factor Contribution* genannt - das wäre präziser und bescheidener gewesen -, dann hätte man dessen Berechtigung wahrscheinlich bald angezweifelt. Heute, nach weltwei-

---

[36] *1. Each row of the factor matrix should contain at least one zero.*
2. If there are m common factors, each column of the factor matrix should have at least m zeros
3. For every pair of columns in the factor matrix, there should be several variables for which entries approach zero in the one column but not in the other
4. For every pair of columns in the factor matrix, a large proportion of the variables should have entries approaching zero in both columns when there are four or more factors
5. For every pair of columns in the factor matrix, there should be only a small number of variables with nonzero entries in both columns.

ter Verbreitung des Zauberworts ‚Simple structure' kann man sich von ihm aus Gründen der Sprachgewalt nicht mehr leicht trennen.

Zum Fortbestand des harten Kerns des SSM haben nicht zuletzt führende Methodiker des Faches beigetragen. Dem Bedarf nach mehr Komplexität bei der faktoriellen Modellierung von Untersuchungsinhalten begegneten sie mit erhöhter Komplizierung ihrer von Anfang an unpassenden Verfahren. Die zurückhaltende Attitüde eines exploratorisch arbeitenden Forschers, der sich für überraschende Befunde offen hält, ist zunehmend der Attitüde eines Variablen-Administrators gewichen, der nach Gutdünken die Einheiten seiner Domäne festlegt und ihnen eigenmächtig Funktionen zuweist, unter denen er die - nur relativ - besten ermittelt.

Für die derzeitige epistemische Lage im Fach ist charakteristisch, dass generell Machbarkeit in der Wertrangordnung einen der ersten Plätze einnimmt. Den Modellen, die man herstellen, manipulieren und der Natur mehr oder weniger auferlegen kann, wird mehr Respekt entgegen gebracht als den Modellen, die der Natur abgewonnen werden. Die Kompliziertheit der Produkte erfolgsstolzer Modellbastler verstellt leicht den Blick dafür, dass sich Entdeckungen von größerer Tragweite in der Regel erst nach ausgiebiger Naturbetrachtung einstellen, bei der mit Zurückhaltung Informationen entgegen genommen werden.

Das Prinzip des SSM kann man mit Lakatos (1977) als Beispiel eines harten Kerns (‚hard core') auffassen, der die Entwicklung konventioneller Forschungsprogramme längerfristig begleitet. Eine Grundüberzeugung wird zu einem ‚hard core', wenn eine Forscher-Community sie für so selbstverständlich hält, dass sie nicht mehr infrage gestellt wird.[37] Für Michell (2000) ist das blinde Festhalten an fundamentalen Voraussetzungen Symptom einer ‚wissenschaftlichen Pathologie': *„A hypothesis is acccepted without a serious attempt being made to test it and this failure of critical inquiry is ignored"* (p. 648).

Ein solches Symptom beschrieb auch Schönemann, indem er die Empörung der ‚Traditionalisten' erwähnte anlässlich der Versuche Bargmanns, das Vorliegen von *Simple Structure* statistisch zu überprüfen: *"[They] scorned it [Bargmann's Test] as a sacrilege of their cherished belief that simple structure is a law of nature"* (Schönemann, 1994, p. 293). Lange vor Lakatos hat Fleck (1935) die wissenssoziologischen Auswüchse der ‚Denkkollektive' mit Falldarstellungen beschrieben. *„Ist ein ausgebautes, geschlossenes Meinungssystem…einmal geformt, so beharrt es beständig gegenüber allem Widersprechenden… Was in das System nicht hineinpasst, bleibt ungesehen oder es wird verschwiegen…oder es*

[37] *„It may be contended that we should cling at all costs to the conception of simple structure, because we have no satisfactory alternative"* (Reybourne & Rath, 1949, p. 127).

wird mittels großer Kraftanstrengung dem Systeme nicht widersprechend erklärt" (p. 35). „Die Beharrungstendenz der Meinungssysteme beweist uns, dass sie ... als stilvolle Gebilde zu betrachten sind. Sie …zeigen als harmonische Ganzheiten besondere Stilmerkmale, die jede einzelne Erkenntnisfunktion bestimmen…[Sie bringen eine ‚Harmonie der Täuschungen' hervor], die dann im Bereiche eines bestimmten Denkstils auf keine Weise aufzulösen ist" (p.45).

Hard core-Voraussetzungen in wissenschaftlichen Forschungsprogrammen führen nach Lakatos auf Dauer zu empirischen ‚Anomalien'. Mit einer wachsenden Zahl von Anomalien geraten die Programme in eine ‚degenerative' Entwicklungsphase. Zur Sicherung des harten Kerns werden dann ‚Schutzgürtel' erstellt und bei Bedrohung verstärkt. Die hochkomplizierten Methoden zur Verarbeitung multivariater Daten, die einseitig mathematisch ausgerichtete Methodiker seit etwa zweieinhalb Jahrzehnten entwickeln, lassen sich als Schutzgürtel verstehen, die das Dogma der SS immunisieren (mehr dazu in Abschnitt 08).

07. Unbeachtete Stimmen der Kritik

Hat man am Prinzip des SSM nicht schon früher Anstoß genommen? Am Prinzip der SSM wurden, was dessen Leistung betrifft, mitunter Zweifel laut, doch seine grundsätzliche Berechtigung wurde so gut wie nie hinterfragt. W. Stephenson (1956) gehört zu den frühesten Zweiflern. Er erkannte, dass der Versuch, Persönlichkeitseigenschaften auf einzelne Faktoren zu reduzieren, auf den Widerstand empirisch vorgefundener Komplexität stößt: *„The problem is to explain complex traits in terms of relatively few primaries" (*p. 7). Mit liberaleren Rotationsmethoden wollte er den von ihm zu Recht so genannten ‚compound traits' einen Freiraum der Repräsentation verschaffen. Doch sein Verfahren war aufwendig, ähnlich wie das Circumplex-Verfahren von Hofstee et al. (1992) und de Raad et al. (1994) (s. Abschnitt 08) mit ihren verwirrenden Einzelergebnissen, die ebenso wenig wie bei Stephenson vom Rotationsideal des SSM abließen. Stephenson favorisierte zeitlebens die faktorielle Q-Technik der Faktorenanalyse, welche ‚Personvariablen' zur Analyse führt, die mit dem von ihm eigens entwickelten Q-Sort-Erhebungsverfahren gewonnen wurden. Die Personvariablen werden interkorreliert und faktorisiert, während sonst im allgemeinen die faktorielle R-Technik eingesetzt wird, bei der man mit Verhaltensvariablen operiert (z. B. Testwerte, Urteilswerte). Stephenson umgeht damit die ungelöste Frage: Welches sind die basalen Merkmale, mit denen die Komplexität des Erlebens und Verhaltens und zugrunde liegende individuelle Dispositionen („compound traits") theoretisch verständlich wären. Zur Klärung dieser Frage ist die exploratorische R-Technik un-

verzichtbar, denn Ergebnisse der Q-Technik müssen durch Konstrukte interpretiert werden, die die Q-Technik selbst nicht zur Verfügung stellt. [38]

Hochgradig unzufrieden mit Thurstones SSM ist auch Guilford (1974, pp.498 f): *"Thurstone's principle of simple structure … is by no means sufficient if we want logical psychological meaning… In numerous instances [in Thurstone's and his students' studies] tests of very different character are often thrown together significantly on the same factors. This does not make good psychological sense…Rarely were varimax factors clean-cut and easy to interpret psychologically…"*. Guilfords Strukturmodell der Intelligenz war eine Reaktion auf das von ihm richtig erkannte Ungenügen der Ergebnisse, das die Anwendung der FA nach konventioneller Vorschrift mit sich brachte. Er sah sich zu zusätzlichen Ordnungsleistungen heraus fordert, die die FA selbst nicht lieferte. Doch setzte er dabei erst am Ergebnis der SS-Transformation an. Den inhärenten Fehler des Thurstone-Prinzips erkannte auch er nicht. Er suchte ihn durch von außen heran getragene Klassifizierungen zu korrigieren.

Ein jüngerer Kritiker, der ähnliche Argumente vorbrachte, ist Allen Yates – Stephenson und Guilford bleiben von ihm indessen unerwähnt – unverständlicherweise. Wie kein anderer vor und nach ihm brachte er die Probleme des SSM auf den Punkt: *"The factors that result from cluster-oriented factor analysis [from blind application of simple structure] are simply an index of success an investigator has had in putting together groups of collinear variables (clusters) - regardless of how complex these same variables might be in terms of their latent determinants. In other words, manifest collinearity among variables is an indication only that they share the same pattern of causal determination; it does not in any way suggest that the shared pattern involves just one latent causal factor"* (Yates, 1987, p. 39). Die Monographie von Yates enthält den konzertiertesten Angriff gegen die ‚Ideologie' des SS-Modells, den ich ausfindig machen konnte. Sein letztes Kapitel beginnt Yates mit: „*Only a radical reorientation of current perspectives will allow researchers to apply exploratory factor analysis in the manner envisioned by its originators as a powerful technique for routine discovery of the underlying bases of observed covariation*" (Yates, 1987, p. 323). Aber auch Yates rührt das SS-Prinzip nicht an, er meint nur,

[38] Eine Relativierung ist hier nötig: Da die *inter*individuelle Varianz von Verhaltensvariablen in einer kurzzeitig geförderten Forschungsperiode relativ leicht feststellbar ist, bedient man sich in der Forschungspraxis vornehmlich der hierfür zuständigen R-Technik. Ungleich aufwendiger wäre die Ermittlung *intra*individuell ablaufender Kovarianz, die größere Zeiträume umfasst. Die *P-Technik*, die für die Aufklärung der dabei anfallenden Daten einzusetzen wäre, ist deshalb in der Forschungsszene eine Rarität, eine gleichzeitige Aufklärung intraindividueller zusammen mit interindividueller Varianz, die auf einer gemeinsamen konzeptuellen Basis stattfinden sollte, wird hauptsächlich aus äußeren (ökonomischen) Gründen selten versucht.

man habe Thurstones ursprünglich liberalere Auffassung ignoriert (ignoriert hat sie der spätere Thurstone allerdings auch selbst). Zur Analyse komplexer Datensätze führt Yates lediglich neue Rotationsalgorithmen ein. Seine Alternativen (*Direct Geomin* und *Direct Geoplane*) wurden wie die Verfahren anderer Autoren, die den unwillkommenen Begleiterscheinungen der SS-Analyse entkommen wollten, überaus kompliziert. Das Gelingen einer Faktorenanalyse wurde von subjektiven Entscheidungen des Forschers abhängig gemacht. Yates' Innovation sollte ein Ausweg aus dem Dilemma werden. Sie blieb von der Fachwelt unbeachtet.

Die SSM-Problematik erkannte auch Rozeboom (1991), der sie für ausweglos hielt: *„For diagnosing the causal grain of common-factor space, rotation to simple structure is so disquietingly fallible that we would surely prefer another criterion were any plausible alternative at hand. (Churchill's aphorism on the inferiority of democracy comes to mind here, namely that democracy is the worst form of government there is - except for all the others…Read 'simple structure' for 'democracy' and 'rotation criterion' for 'form of goverernment')*" (p. 587). Rozebooms HYBALL Rotation, bei der mehrere Koordinaten (Unterräume) anstelle einzelner Koordinaten sukzessive rotiert werden („somewhat more holistic than a simple sequence of planar rotations", p. 587) erweist sich als ein ähnlich fragwürdiger Kompromiss wie der von Yates.

Zu den wenigen Kritikern, die das SSM grundsätzlicher infrage stellen, gehören auch Schönemann & Borg (1996), die resigniert konstatieren, *„dass das Einfachstruktur-Kriterium (formuliert als Varimax-Kriterium) in der faktorenanalytischen Praxis routinemäßig angewendet [wird]"*. Sie ergänzen mit kritischerem Blick: *„Auf der Strecke bleibt dabei ‚die wichtige Frage, warum man Einfachstruktur in vielen Systemen von Variablen überhaupt erwarten sollte"* (Steiger ,1994, p. 204). Im Grunde sei „die Einfachstrukturhypothese wenig plausibel…". Sie behaupte, Testaufgaben mit Nicht-Null-Ladungen auf allen Faktoren könne es „unmöglich" geben, damit „spannt [sie] den Wagen vor das Pferd" (nach Guttman, 1992, p. 186).

Den Typ von Irrtum, den sich die Proponenten der SS-Analyse leisten, hat Gigerenzer (1978) treffend diagnostiziert – allerdings an einem anderen Beispiel dimensionsanalytischer Modellbildung. Unkritische Modell-Anwender in der Psychologie ignorieren, dass ihre Modelle mit dem psychologischen Gegenstandsbereich durch inhaltliche Vorannahmen verknüpft sind, die selbst nicht unmittelbar geprüft werden. Ignoriert wird dabei das, was Gigerenzer eine ‚Implikationsthese' nennt: *„[Diese These] beinhaltet, dass jedes mathematische System (z. B. eine dimensionsanalytische Methode) bereits durch seine Anwendung auf einen psychologischen Gegenstandsbereich G eine psychologische Theorie über diesen Gegenstandsbereich impliziert."* (p. 110). [39] Das for-

[39] Mit der Implikationsthese Gigerenzers wurde die„central thesis" einer Publikation zur multidimensionalen Skalierung von Smith & Jones (1975) vorweg genommen: *„All data analysis and all scaling involve fundamental assumptions about the*

male Relationssystem (FRS) ist mit dem ‚empirischen Relationssystem' (ERS) untrennbar verschränkt (Gigerenzer verweist hier auf eine Unterstützung durch Suppes & Zinnes, 1963). Fallen die Systeme FRS und ERS in entscheidenden Punkten auseinander, entstehen ‚Divergenzartefakte' mit ‚theoretisch wertlosen Ergebnissen' (p. 111).

Auf den vorliegenden Kontext angewandt: Das FRS der Einfachstruktur setzt bei seiner Anwendung ungeprüft voraus, dass den auf der ERS-Ebene zugänglichen Variablen solitäre Varianzquellen zugrunde liegen. Gigerenzer fordert: *„Um ein Untersuchungsergebnis gegen eine Interpretation als Divergenz-Artefakt abzusichern, muss der Forscher … das durch das mathematische Modell implizierte gegenstandsbezogene psychologische Modell explizieren …"* (p. 116). Genau dies haben Faktorenanalytiker bislang versäumt. Sie glauben, ihre mit mathematischer Präzision auf empirische Daten angesetzten Analysen würden ihnen die Struktur der psychologischen Realität automatisch ausliefern. Zusätzliche und unabhängige Denkmittel einzusetzen, wird für überflüssig und sogar anrüchig gehalten, da sie mit dem Makel des ‚Subjektiven' behaftet werden.

08. Führen nichtfaktorielle Verfahren weiter?

- Was leisten Circumplex-Verfahren?

Mit vielerlei Maßnahmen haben Methodiker auf die unbefriedigenden EFA-Resultate reagiert. Mathematische Überbau-Gebilde sollten die Symptome kurieren (Revenstorff, 1980, p.12). Den Mehrfachladungen bei Variablen (z. B. Fragebogen-Items), die Abweichungen vom SS-Ideal darstellen und in der faktoriellen Persönlichkeitsforschung ein alltägliches Ärgernis sind, versuchte man durch aufwendige Zusatzberechnungen gerecht zu werden. Dafür gaben Hofstee et al. (1992) mit ihrem AB5C-Modell (Abridged Big Five Dimensional Circumplex) ein Beispiel ab. Das SS-Resultat adjektivischer Person-Beschreibungen suchten sie mit der Circumplex-Methode von Wiggins (1979) aufzubessern. Die Circumplex-Methode lässt das Zusammenvorkommen von jeweils zwei faktoriellen Traits beim gleichen Item zur Abbildung zu, es entlastet somit von den Restriktionen des SSM nur partiell.

Aber dies ist wieder ein hybrider Kompromiss, das Verfahren *„does not deal adequately with those …variables that load highly on more than two … factors"*. Circumplex

---

psychological processes that lead to the data and the scaling solutions under consideration. In particular, current scaling methods, in our view, should be regarded with deepest suspicion, precisely because they are based on doubtful or untested psychological assumptions" (p. 44).

macht also nur Teil-Zugeständnisse an die Komplexität der Realität und ist im Ergebnis weder „definitive nor comprehensive“ (Hofstee et al., p, 161). Am SS-Prinzip selbst wird nicht gerüttelt, das Dimensionsproblem bleibt unangestatet. Die Autoren trauen sich nur zu „[to] propose a partial liberalization of simple structure“(p. 147), wo doch erst eine Abschaffung des SS-Prinzips das Komplexitätsproblem lösen würde. Dass Circumplex-Methoden versagen, wenn „all items are best described by more than two factors”, stellen auch Acton & Revelle (2004, p. 26) fest, die die psychometrischen Kriterien für das Vorgehen mit Circumplex explizieren. Die Voraussetzungen für eine Anwendung von Circumplex-Methoden sind im Übrigen sehr anspruchsvoll – die Autoren führen zehn Voraussetzungen auf –, weshalb Circumplex-Verfahren selbst dann kaum zu empfehlen sind, wenn nur zweifaktoriell beschreibbare Datensätze zur Analyse anstehen.

- Sind konfirmatorische Verfahren eine Alternative?

Die konfirmatorischen Verfahren (CFA), die durch Strukturgleichungsmodellierung (SEM) auf ein flexibleres Niveau gebracht wurden, scheinen mit ihren schwierigen Algorithmen die primär inhaltsorientierten Forscher beeindruckt und die exploratorischen Verfahren weitgehend verdrängt zu haben. Anstatt als EFA-Forscher mit latenten Variablen zu rechnen, die man nicht unbedingt vorauszusagen wagt, ist es Mode geworden, diese nach Gutdünken zu erfinden, um sie anschließend mit konfirmatorischer Technik zur empirischen Realität ins Verhältnis zu setzen. Zu diesem Zweck werden, meist mit Trial-and-error-Blindheit, Unmengen von Passungsversuchen durchgeführt.

Doch orientieren sich Strukturgleichungsmodelle ebenfalls durchweg am SS-Prinzip. Somit bleibt auch auf diesem Wege die Komplexität latenter Varianzquellen verdeckt.[40] *„Die neueren Strukturgleichungsmodelle – zunächst, wie gehabt, euphorisch gefeiert … - tragen die Probleme der Faktorenanalyse in verschärfter Form fort“*(Schönemann & Borg, 1996, p. 241). Fast immer verfehlen sie die zugrunde liegenden Strukturen.[41]

---

40 Basilevsky (1994, p. 415) beschreibt die Tendenz in der CFA-Praxis zum Herauspräparieren singulärer Faktoren bei den jeweils untersuchten Variablen u. a. wie folgt:*„…we may wish to impose zero restrictions on the loadings. Values other than zeroes can also be used, but zeroes are most common in practice”*.

41 *„Dass keine der latenten Ursachen eindeutig definiert ist, wird im LISREL-Manual taktvoll verschwiegen“* (Schönemann & Borg, 1996, p. 250). Das „Unbestimmbarkeitsproblem“, ein rein mathematisch-formales Problem, sei im LISREL-Fall „tatsächlich verschlimmert …, weil man dort meist wesentlich mehr latente Variablen postuliert als im Modell der multiplen Faktorenanalyse…“ (p. 250).

Die Ergebnisbilanz der SEM-Forschung fällt entsprechend mager aus („poor results", Beauducel & Wittmann, 2005, p. 42). Man ist dabei, sich mit den unzureichenden Ergebnissen auch dieser Forschung abzufinden. Es bleibt bei Feststellungen wie *„the assumption of simple structure is probably a typical… simplification bias"*, aber leider ‚necessary' (Beauducel & Wittmann, 2005, p. 43). *„…Simple structure models of personality are unlikely to meet conventional or even fairly relaxed goodness-of-fit criteria…Overemphasis on simple structure … may explain some of these problems"* (p. 44). In einem kritischen Review-Artikel wiesen Mac Callum & Austin (2000) auf die Probleme der konfirmatorischen Verfahren und auf den ‚confirmation bias' ihrer Anwender hin, die sich mit *Goodness-of-fit*-Werten und willkürlich gewählten Kriterien begnügen, welche nur scheinbare Passungen herbeiführen. Nach Ansicht der Autoren sollten Analyse-Ergebnisse nach sachorientierten und nicht primär nach formalen Maßstäben bewertet werden.[42] Wenn die überprüften Modelle das Fit-Kriterium nicht erreichen, werde dies oft leichtfertig ignoriert („working with imperfect models", Titel eines Artikels, MacCallum, 2003). Oder es werde nicht bedacht, dass die Passung des selektierten Modells von anderen nicht-geprüften Modellen leicht übertroffen werden könnte. Schon Kaiser hat von der CFA nicht viel gehalten: *„I cannot resist saying that, for me at least, the earlier exploratory thrashing about was much more fun - and perhaps even represented more progress - than the forth-coming confirmatory prettying-up"* (Kaiser, 1970, p. 406). –

Ernüchternde Ergebnisse aus weiteren Vergleichen zwischen CFA und konventioneller EFA findet man bei Church & Burke (1994) und Ferrando & Lorenzo-Seva (2000). Die Kritik von Cliff (1983) liegt auf der gleichen Linie. Eine jüngste umfangreiche CFA-Studie mit Simulationsdaten kommt zur Schlussfolgerung: *„…trait models [of personality] assuming simple structure tend to be rejected with CFA…". „…there will always be some small distortion of simple structure"* (Beauducel & Wittmann, 2005, p. 72). Mit Anwendung des SSM bei Datensätzen, insbesondere bei Persönlichkeitsdaten, die der CFA unterworfen werden, wird eine Kluft (*'gap'*) geschaffen *„between the large body of results based on exploratory factor analysis and CFA in personality psychology"* (p. 73). Die exploratorische Analyse der konventionellen Art, die hier gemeint ist, weist trotz Ausrichtung auf das SSM Sekundärladungen bei vielen Variablen auf, die beim blinden Vorweg-Verteilen von Faktorgewichten im konfirmatorischen Prozedere nicht erahnt werden („*…there is no [prior] knowledge of secondary loadings*" , p. 43). LISREL-Ergebnisse hält Vittadini (1989) in jedem Fall für indeterminiert, weil die konstruierten latenten Modellvariablen von manifesten

[42] *„Die praktisch schrankenlose Plastizität des LISREL-Modells … unterhöhlt nicht nur den Anspruch auf statistische Inferenz, sie fordert fast zum Missbrauch auf, denn bei hinreichender Geduld wird sich schon irgendein Kausalmodell finden, das für die gerade vorliegenden Daten nicht abgelehnt werden muss."* (Schönemann & Borg, 1996, p. 250).

Variablen abhängig gemacht werden: *„one may actually be confirming the model because the manifest variables are determined by other variables than those hypothesized, which happen to have the same pattern of relationship to the manifest variables as given by one's hypothesis…One can never regard structural hypothesis as true as opposed to 'confirmed'* "(p. 428).

Konfirmatorische Verfahren können sich dann als nützlich erweisen, wenn die latenten Variablen in der jeweiligen Domäne zuvor durch exploratorische Analysen gesichert wurden. Velicer & Jacksons (1990) Argumentation geht in diese Richtung: *„Exploratory analytic approaches … should be preferred except for those cases where a well-defined theory exists. Exploratory approaches avoid a confirmation bias, do not force a theory-oriented approach prematurely, and represent a conservative strategy"* (p. 21). Die Frage, ob zur Überprüfung eines exploratorisch gewonnenen und theoretisch befriedigenden Ergebnisses eine konfirmatorische Modellrechnung überhaupt noch sinnvoll einzusetzen ist – es käme doch einer Replikation des exploratorischen Ergebnisses gleich – , müsste man wohl verneinen.[43]

- Wie sind mathematische Erkenntnismittel generell zu nutzen?

Die im gegenwärtigen Wissenschaftsbetrieb zunehmende Neigung, den mathematischen Techniken blind zu vertrauen und ihren Schöpfern die Verantwortung für deren Anwendung zu überlassen, verdient generell Kritik. In den Schriften selbstkritischer Methodologen finden sich Warnungen: „*Those who firmly believe that rigorous science must consist largely of mathematics and statistics have something to unlearn. Such a belief implies the emasculation of the basic substantive nature of science. Mathematics is content-less, and hence not – in itself – empirical science… rigorous treatment of content or subject matter is needed before some mathematics can be thought of as a possibly useful (but limited) partner of empirical science*" (Guttman, 1971, p. 42). [44] Mit ähnlicher Stoßrichtung

---

[43] Rost (2003): *„Der Modellspezialist [Spezialist im Modellieren mit Strukturgleichungen, log-linearen Modellen, Item-Response-Modellen usw.] …kann mit Daten überhaupt nichts anfangen, wenn man ihm nicht sagt, welche Variable mit welcher anderen interagieren soll…, welche latenten Variablen es geben soll … usw."* Mein Kommentar hierzu: Der heutige ‚Modellspezialist' traut sich a priori zu, was der exploratorisch arbeitende Faktorenanalytiker früherer Jahrzehnte trotz Analyse massenhaft vorliegender Daten nicht einmal im Nachhinein hat gewinnen können, nämlich eine approximative Einsicht in die latenten Parameter menschlichen Denkens, Fühlens und Verhaltens. Diese Parameter, so setzt man heute vielfach voraus, könne man sich ausdenken, man müsse sie nur mit Versuch-und-Irrtum-Strategie durch Modellfitting herausschütteln.

[44] Das Zitat ist einem Artikel von Barrett (2002) entnommen, der die konventionelle Psychometrie insgesamt kritisiert, weil sie durch Mess-Operationen die

wird der Skeptiker Guttman von Schönemann zitiert: *„There remains the danger of seeking data merely to fit axioms“, dazu der Kommentar: “In hindsight, these warnings sound positively prophetic in anticipating the present malaise in mathematical psychology some 20 years before Cliff (1992) noticed it…“.* Schönemann (1994, p. 294) nennt noch Narens & Luce (1993) als Kritiker der Malaise. Ein Pionier der Mathematischen Psychologie, William K. Estes, beklagte schon 1975 die Situation seiner Spezialisierung: „… *it is clear that many investigators in our field are not entirely happy with their current situation“* (p. 263). Er bedauert die Kluft zwischen der mathematischen und inhaltlich-theoretisch ausgerichteten Psychologie und zitiert, um eigene Revisionsvorschläge zu unterstreichen, Leont'ev & Dzafarov (1973, p. 20): *„An analysis of the present situation shows that contemporary psychology and contemporary mathematical instruments are still not compatible enough with one another to allow mathematization to assume a central place in the development of psychological knowledge; the reason for this is not only the low level of sophistication of the latter … What is required is a continual interaction between mathematics and psychology, an interaction that … would lead to a revision of existing mathematical methods into forms more amenable to the proposed mathematized conceptual systems.“*

Zugang zur psychologischen Inhaltsebene gewinnt man nicht primär durch formale Modelle, sondern durch die Gesamtheit der Erfahrungen, die zur jeweiligen Domäne gehören. Das durch frühere Erfahrungen gewonnene Wissen und Vermuten wird durch Forschung präzisiert und geprüft, alltagsprachlich kommunizierbare Repräsentationen spielen dabei eine große Rolle. Auch vor allem an diesen sollte, wenn möglich, die Passung der formalen Modellierung psychologischer Sachverhalte bemessen werden. Dazu haben sich kritische Autoren wiederholt geäußert:

- *„…Neither algorithmic sophistication, nor axiomatic rigor alone are apt to advance our knowledge much if they are cultivated in an empirical vacuum“* (Schönemann, 1981, p. *412*).
- *May I … insist once again on the absurdity of divorcing the mathematical or statistical evidence from evidence procured by other means? … The sole claim of mathematical analysis should be to verify, by appropriate calculation, the hypotheses commonly advanced on the basis of much broader and more general lines of evidence“* (Burt,1949, p. 107).
- *„Let us try to be free of …a priori mathematical and statistical considerations and prescriptions – especially codes of permission. Instead, let us try to think substantively*

Inhaltsebene (Intelligenz, Persönlichkeit) entleert. Er plädiert, gewissermaßen ersatzweise, für Offenheit gegenüber Anwendungsfragen. Barrett propagiert eine sogenannte „applied numerics“, die von fragwürdigen theoretischen Erwartungen befreien und pragmatisch umso mehr nützen soll (Absage an theoretische Ansprüche).

… and focus directly on the specific universe of observations with which we wish to do our business. “ (Guttman, L., 1971, p. 346).

- *„Factorists with more mathematical training than the rest of us have been addressing themselves to problems…on a technical rather than upon a fundamental level…In most cases [their models are] irrelevant …to the problems …of those for whom factor analysis is a research tool…* “ (Butler, J. M., 1969, 252-3).
- *„These techniques [“for rotating factors into 'psychologically meaningful‘ positions”] …became the stock in trade of practicing factor analysts. …It is probable that future historians will be severely critical of them and of their users; critical of the techniques because of ... their users for extravagant claims on their behalf.* “ (Maxwell, 1959, p. 228).

Angesichts der zunehmend formalistischen Ausrichtung der psychologischen Forschung (und leider auch der Lehre) wären die nicht-mathematischen Erkenntnismittel, die oft als ‚hermeneutische‘ und ‚phänomenologische‘ mit negativem Vorzeichen abgedrängt werden, kompensatorisch besonders zu pflegen und zu fördern. Wenn sie tiefgreifender operieren (‚on a fundamental level‘, Butler) und zukünftig zur kritischen Rückschau auf die ‚extravagant claims‘ der Formalisten benötigt werden (Maxwell), müsste man sie schon heute mit Nachdruck am Leben erhalten. Innerhalb der Gesamtheit unserer Erkenntnismittel erfüllen die mathematischen doch nur einen Teil unserer Wünsche.[45] Präzision im Detail und die zu diesem Zweck eingesetzte ingeniöse Mathematik und Statististik sind nützlich, aber nicht für jede menschliche Geistesregung unentbehrlich. Sie können schädlich sein, wenn sie unnötigerweise oder im Übermaß eingesetzt werden – das zeigt das Debakel mit SSM –, während man die basaleren, umfassenderen, holistischer ausgerichteten, wenn auch zeitweise weniger scharfen Erkenntnismittel, vielleicht

[45] Wenn die vernachlässigten Vorgehensweisen *phänomenologische* genannt werden, erscheinen sie möglicherweise wissenschaftshistorisch belastet. Auch mit der Bezeichnung *hermeneutisch* könnte man Assoziationen zu geronnenen Denkstilen hervorrufen und von Meinungsgegnern entsprechend abgestempelt werden. Als Desiderat darf man indessen einbringen, dass in allen Phasen der Problemlösung irrationale Hindernisse überwunden werden sollten, die sich der Verwendung menschlicher Erkenntnis mit ihren weit reichenden Verzweigungen entgegen stellen. Die Bemühungen in den Natur-, Geistes- und Kulturwissenschaften sind in ihrer Gesamtheit Niederschlag einer Entfaltung epistemischer Möglichkeiten. Sie sollten bei Bedarf in jeder einzelnen Disziplin, wenn auch mit unterschiedlichen Akzenten, immer abruffähig bleiben. Dies ist auch das Hauptanliegen des Dissidenten Sigmund Koch, 1999, und anderer ungehörter Rufer, die die ständig ins Objektivistische abgleitende Wissenschaft kritisieren (Bridgman, 1959).

ungewollt, aber doch systematisch ins Abseits drängt.[46] Für ein holistisches Vorgehen in der Forschung, das die methodisch isolierten Beobachtungen auf Vereinbarkeit mit dem Hintergrundwissen prüft, plädiert auf wissenschaftstheoretischer Basis auch Gawronski (2000).

Diskussion zu Kapitel 1 und Ausblick

Das vorliegende Kapitel ist keine Absage an die Faktorenanalyse überhaupt, sondern nur an eine Anwendung, die ihre Voraussetzungen ignoriert. Gigerenzer & Strube (1978) propagierten eine ‚annahmenkritische' Anwendung der verfügbaren Verfahren. Hätte man die von diesen Autoren geforderte Kritik der Annahmen vorgenommen, auf welche sich Faktorenanalytiker stillschweigend stützen, hätte schon früher das erreicht werden können, was sich Yates erhofft hatte. Dieser Autor hatte einen Paradigmenwechsel für notwendig gehalten, [47] die ‚Pathologie' der faktoranalytischen Forschung wäre sonst nicht zu beheben und dem ‚Morast' ihrer bisherigen Ergebnisse nicht zu entkommen.

Zuallererst wäre die ‚Ritualhandlung' des ‚Little Jiffy' einzustellen, die von Gigerenzer & Strube (p. 81) angeprangert wird (Little Jiffy = rezeptartige Anwendung der FA, die mit einer Varimax-Rotation endet). Der Geist der Kritik wäre sodann wieder aufzunehmen, der im deutschsprachigen Raum auf dem Münchener Symposium mit Kallina (1967), Kalveram (1970), Gigerenzer & Strube (1978) und Revenstorf (1980) auflebte, dann aber versiegte. Das Komplexstruktur-

---

[46] S. Jevons (1873): *„Ich neige dazu, an mathematischen Autoren herumzunörgeln, weil sie oftmals bejubeln, was sie alles leisten können, und es unterlassen, darauf hinzuweisen, dass das, was sie machen, nur ein winzig kleiner Teil dessen ist, was gemacht werden könnte. Sie zeigen…allgemein die Neigung, die Existenz von widerspenstigen Problemen noch nicht einmal zu erwähnen …"* (zitiert nach Rescher, 1985, p. 124).

[47] Die derzeit tonangebenden formal-modellierenden Fachvertreter scheinen einen ‚Paradigmenwechsel' kategorisch auszuschließen. So spricht etwa Rost (2003) von „ehernen Gesetzen, die jede Modeströmung überstehen… oder gar gestählt aus ihr hervorgehen". Nach Rost können Moden und Zeitgeist der Psychologie den Methodeneinsatz zwar am Rande affizieren, doch dies werde nicht dazu führen, *„dass sich das bisherige Arsenal an Forschungsmethoden als falsch oder unbrauchbar erweist. Vielmehr führen sie [die Moden] nur dazu, dass sich das Methodenarsenal um wesentliche Teile erweitert und bereichert."* Dass auch fehlkonstruierte Teile in diesem Methodenarsenal vorhanden sein und großen Schaden anrichten können und dann grundlegend überarbeitet werden müssen, hätte Rost in seiner generalisierenden Rückschau eigentlich zu ergänzen.

Paradigma braucht zu seiner Bestätigung ebenso wie zu seiner Ablehnung – falls sich eine solche mit hier vielleicht übersehenen Gegengründen durchsetzen möchte – ein aktives Engagement. Empfohlen wird sodann der umfassende Einsatz des im nächsten Kapitel vorgeschlagenen alternativen Verfahrens. Strube (2000, p. 117) übte zurecht an denjenigen Kritik, die bislang mit dem ‚senselessly adhering to the ususal methods…' nur unzufrieden waren, ohne dem etwas entgegen zu setzen: „*Most of [existing work] discuss negative examples without giving positive ones*"(p. 117). Mit Anwendung der Faktorenanalyse, die sich vom Komplexstruktur-Modell leiten lässt, sollte eine Befreiung vom Dogma Thurstones und – vielleicht nach weiteren notwendigen Verfeinerungen der Methode – eine optimale Struktur-Abbildung der latenten Varianzquellen manifester Observablen erreicht werden.

Henry F. Kaiser, der für das Little Jiffy die Voraussetzungen schuf, wäre zu einer radikal neuen Ausrichtung der Faktorentransformation sicherlich bereit gewesen – er ist zu früh verstorben (1927-1992). Denn er hatte sich am Ende eines Vortrags über die ‚Second generation Little Jiffy' vor der Psychometric Society für solche Fälle ausdrücklich angeboten: *"For the future, I can assure you of one thing: if any of you folk… come up with some Big Breakthroughs I shall be waiting in the wings ready and eager to paste them together to produce the next generation Little Jiffy"* (Kaiser, 1970, p. 415). Gene Glass (1991) charakterisiert Kaiser in einem Nachruf als ‚respektlos', eine Eigenschaft, die Kaiser offenbar nicht übel anstand, denn, so Glass weiter, *"irreverence must be a necessary ingredient in the recipe for creativity. Whoever worships received wisdom too ardently will never see beyond it."* Das komplexitätsmodellierende Verfahren könnte 'respektlos' erscheinen, da es Kaisers Varimax auf den Kopf stellt. Doch hätte Kaiser meine Unbotmäßigkeit wahrscheinlich eher begrüßt als abgelehnt, so wie er Resperktlosigkeiten selbst praktizierte und von den Studenten seiner – und sicher auch der nächsten Generation – erwartete.

Kapitel 2

Komplexe Strukturen aufspüren. Faktorenanalysen mit Variminrotation

Vorspann

Empirie macht sich nirgends stärker geltend als dort, wo sie unsere Erwartungen durchkreuzt und uns eines Besseren belehrt. Als ich zum ersten Mal auf den Gedanken kam, das SS-Prinzip durch Anwendung des Rotationsverfahrens Varimin in sein Gegenteil zu verkehren, konnte ich nur hoffen, dass meine Erwartungen nicht enttäuscht wurden. Ich konnte nur wünschen, dass mit der Anwendung des neuen Verfahrens komplexe Strukturen der analysierten Variablen sichtbar würden und dass auf diesem Wege die Realität der untersuchten Domänen angemesse-ner modelliert werden würde als mit der konventionellen Technik des Simple Structure-Modelling (SSM).

Meine Unsicherheit war begründet. Variminstrukturen konnten von den Initialstrukturen, die die Faktorenanalyse standardmäßig liefert und die ohne ein weiteres Zutun komplex schon sind, nicht allzu verschieden sein. Hätte man nicht in

den Jahrzehnten faktorenanalytischer Praxis die Vorteile schon erkennen müssen, die initiale Faktorenlösungen mit sich bringen, wenn denn die initiale Komplexität einer Faktorenstruktur tatsächlich schon Vorteile mit sich bringen oder ankündigen sollte. Vielleicht werden Modelle der Wirklichkeit durch Zunahme an formaler Komplexität, die durch eine Varimin-Rotation bewirkt wird, sogar verfehlt. Ohne genaue Kenntnis dessen, worauf das Complex Structure Modelling (CSM) hinaus läuft, ist das nicht auszuschließen.

Zu den willkommenen Befunden, die im zweiten Kapitel dargestellt werden, gehört, dass Variminrotationen der Faktoren dann, wenn sie die Initialstruktur verändern, den Einblick in die latenten Bedingungen der jeweils untersuchten Domäne regelmäßig verbessern, sogar oft erst ermöglichen. Varimin verändert die Initialstruktur mehr oder weniger erheblich, wenn drei oder mehr substantielle Faktoren extrahiert wurden.

Wie aber lassen sich die bei komplexen Lösungen resultierenden Faktoren interpretieren? Nach Thurstone und Nachfolgern sollen Faktoren erst nach einer SS-Transformation interpretierbar sein. Nicht bedacht hatten die Faktorenanalytiker der ersten Stunde indessen eine bahnbrechende methodische Überlegung, die in der Linguistik, speziell in der Phonologie (1952), ausgearbeitet wurde und die dem so genannten ‚Minimalpaarvergleich' den Weg ebnete. Zwei Variablen A und B mögen beide komplex sein, weil jede von ihnen eine Vielzahl latenter Merkmale aufweist. Diese lassen sich in ihrer Gesamtheit nicht erfassen. Wenn aber (n – 1) Merkmale von A und B eine nahezu gleiche Ausprägung haben und nur bei einem Merkmal gegensätzliche Ausprägungen auftreten, dann darf man sich auf die Deutung des Unterschieds zwischen A und B, wenn ein solcher tatsächlich vorliegt, auf dieses eine Merkmal beschränken (s. Tabelle 2.01). Gleich ausgeprägte Merkmale von A und B (a, b, c, e) brauchen nicht beachtet zu werden, wenn es nur um die Identifikation eines Kontrast-hervorrufenden Merkmals (d) geht. Stuhl und Schemel unterscheiden sich nur durch das Vorhandensein oder

Tabelle 2.01

Verdeutlichung eines Minimal-Paar Falles: Die Variablen A und B zeigen beim Merkmal d kontrastierende Ausprägungen

	Merkmale				
Variablen	A	b	c	d	e
A	-2	+1	-1	**+3**	+2
B	-2	+1	-1	**-2**	+2

Nichtvorhandensein einer Rückenlehne, Stute und Hengst nur durch das Geschlecht, Berg und Hügel nur durch die Größe einer Erderhebung. Der Aufweis einzelner kontrastierender Merkmale durch Minimalpaarvergleiche, wie sie mit CSM möglich werden, ist leichter und verlässlicher als das Suchen nach einer Gemeinsamkeit bei Variablen, die nach Anwendung von SSM durch einen Faktor in einen Topf geworfen werden. Die Gemeinsamkeit der SSM-faktoriell geclusterten Variablen kann sich nur auf gemeinsame zugrunde liegende Merkmale stützen. Diese aber bleiben mit Anwendung der SSM-Methode verdeckt. Man beruft sich deshalb im Einzelfall auf ein hypothetisch gemeinsames „faktorielles", also auf ein ad hoc zu bildendes Konstrukt. Dieses ist in der Regel als solches nicht schon naheliegend, wenn überhaupt konzeptuell existent, jedenfalls ist es nicht „simple". Warum haben Methodologen die naheliegende Prüfstrategie des Minimalpaar-Vergleichs bei der Evaluation ihrer Verfahren nie genutzt? Weitere Vorteile einer Faktorenanalyse mit komplextheoretischer Ausrichtung werden in den zehn beispielhaften Analysen des zweiten Kapitels demonstriert.

Ausgangslage und Zielsetzung

In Kapitel 1 wurde eine Kritik an der konventionellen exploratorischen Faktorenanalyse begründet. Die manifesten Variablen einer empirischen Domäne, so wurde argumentiert, resultieren generell aus *multiplen* Varianz- bzw. Kovarianzquellen. Dass einzelnen manifesten Variablen nur jeweils eine Bedingung zugrunde liegt, wie dies das SS-Prinzip Thurstones voraussetzt, darf in der Regel nicht vorausgesetzt werden. Das Prinzip wurde als ein fehlkonzipiertes verworfen.

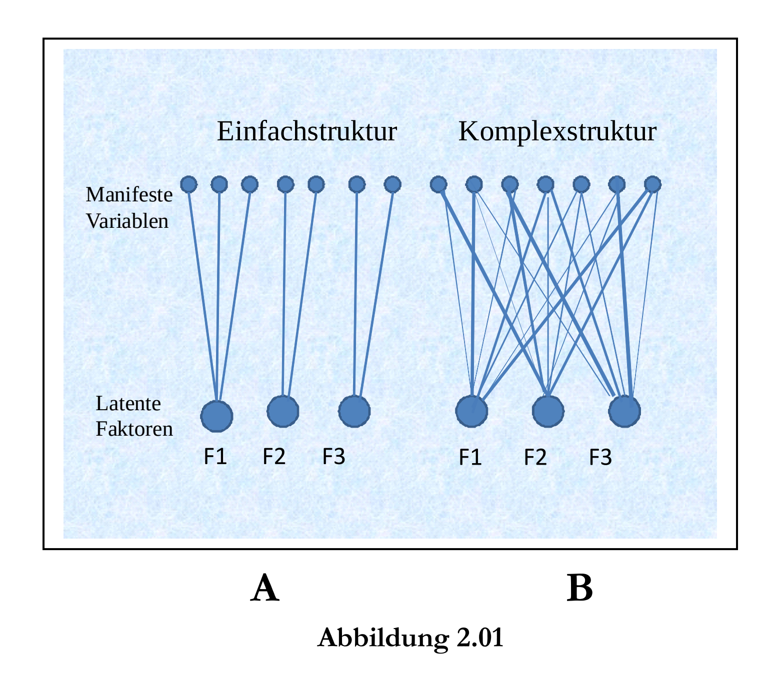


Abbildung 2.01

Einfachstruktur (A) und Komplexstruktur (B) der Beziehungen zwischen latenten Kovarianzquellen und manifesten Variablen

Die Abbildungen 2.01A und 2.01B veranschaulichen den Unterschied zwischen Einfach- und Komplexstruktur. In beiden Fällen ist auszugehen von der Ebene manifester Variablen, die interkorreliert werden mit der Absicht, die der Korrelationsmatrix zugrunde liegenden Faktoren aufzuklären. Faktoren sind die Quellen der Kovarianz der manifesten Variablen. Das Simple Structure-Modell (SSM) Thurstones unterscheidet sich vom hier befürworteten Complex Structure-Modell (CSM) in Folgendem: Eine Rotation der Faktoren zur SS (Abbildung 2.01A) ord-

net einzelnen manifesten Variablen jeweils nur einen extrahierten Faktor zu bzw. so wenig Faktoren wie möglich. Eine Rotation zur CS dagegen (Abbildung 2.01B) ist darauf aus, individuelle Variablen mit möglichst vielen extrahierten Faktoren hypothetisch zu verknüpfen. Nach dieser Konzeption wird angenommen, dass die meisten Faktoren in der Regel zur Kovarianz einzelner Variablen einen – je nach Variable verschieden großen – Beitrag leisten.

Unter den Transformationen, die zur SS führen sollen, werden im Forschungsalltag meist orthogonale Verfahren gewählt, oblique Verfahren nur selten, und von den orthogonalen Verfahren wird fast ausschließlich Kaisers Varimax-Rotation eingesetzt (s. Abbildung 2.01).

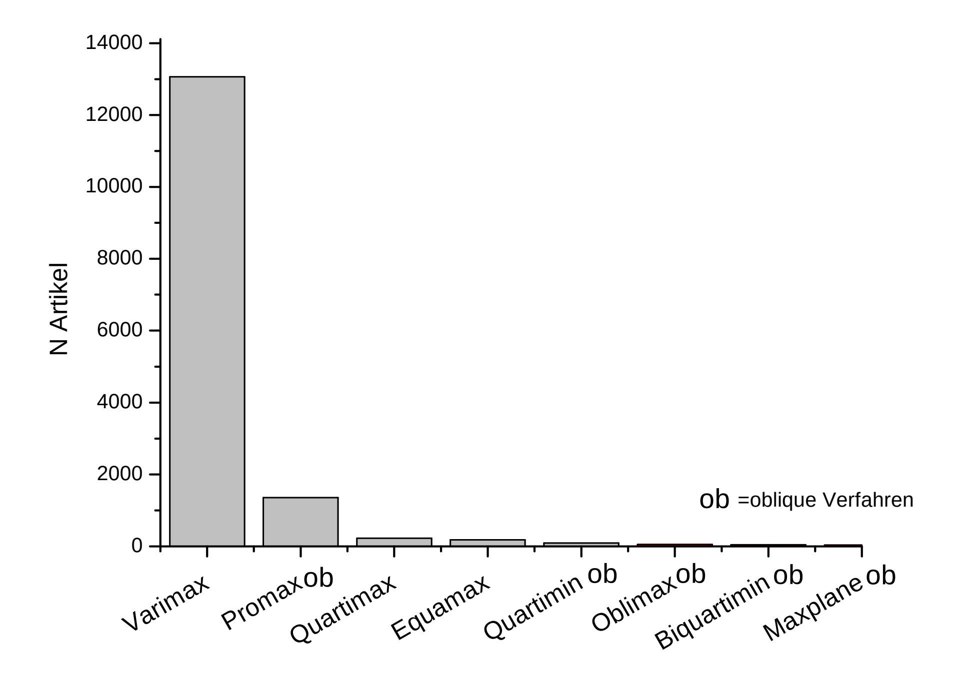


Abbildung 2.02

Anzahl von Artikeln, in denen auf Faktorenrotation der Einfachstruktur (SS)verwiesen wird. Ergebnis einer Internet-Suche nach Rotationsbezeichnungen in SCIENCE DIRECT (Elsevier, 2008)

Nach den Überlegungen von Kapitel 1 sind die Rotationsverfahren, die eine SS-Struktur modellieren wollen (Abbildung 2.02), generell fehlkonzipiert. Das SSM verhüllt die in der Initiallösung oft schon erkennbare Komplexität latenter Bedingungen, anstatt diese als hinzunehmende Realität zu betrachten und zu optimieren. Dabei ist anzumerken, dass die Initiallösung ein befriedigendes Komplexitätsmodell nur beim Sonderfall zweifaktorieller Lösungen schon gewährleistet. Bei Datensätzen mit mehr als zwei Faktoren wird mit der Initiallösung die Komplexität der Kovarianzquellen analysierter Variablen in der Regel nicht optimal manifest. Der Grund ist – und das ist wichtig –, dass bei der Extraktion eines Faktors die Varianz, die die Matrix der Korrelationen – bzw. *nach* einer Faktorenextraktion die jeweilige Restmatrix – enthält, für den betreffenden Faktor überoptimal ausgeschöpft wird. Dies führt zu einem mehr oder weniger großen Überhang an Kommunalität bei den früher extrahierten Faktoren und zu einer entsprechenden Benachteiligung bei den später extrahierten Faktoren (Gorsuch, 1974, p. 161). [48] Zur optimalen Manifestation der Kovarianzquellen wird deshalb eine Faktorenrotation durchaus auch für notwendig gehalten. Doch soll diese das Gegenteil von dem leisten, was eine Simple Structure-Rotation zum Ergebnis hat.

Um dies zu erreichen, habe ich, wie in Kapitel 1 erwähnt, ein Rotationsverfahren konzipiert, das ich *Varimin* nannte, weil es eine Umkehrform von Varimax darstellt. Mit Varimax wird durch paarweises Drehen der Faktor-Koordinaten die Varianz der quadrierten Faktorladungen pro Faktor *gesteigert*. Der Vorgang wird durch Iterationen so lange wiederholt, bis sich die Summe der Ladungsvarianzen für die Gesamtheit der Faktoren nicht mehr steigern lässt. Das Kriterium V (s. Gleichung 1), das Varimax-Kriterium, wird maximiert.

---

[48] Die Faktorenextraktion führt dazu, dass die Interkorrelationen der Initialfaktoren bzw. die Kongruenzindices Phi (nach Tucker), mit denen man ihre Ähnlichkeit untereinander ermittelt, alle exakt Null ergeben. Die Faktoren der Initiallösung haben also untereinander eine maximal orthogonale Beziehung. Doch führt der mathematische ‚Zwang' zur Nullkongruenz unter den Faktoren, die auf die jeweils analysierten Gegenstände oder Sachverhalte ‚keine Rücksicht nimmt', meist zu einer gewissen Verzerrung bei der Modellierung empirischer Verhältnisse. Die Variminlösung korrigiert Abweichungen vom Optimum einer realitätsgetreuen Modellierung, die initiale Faktorenstruktur wird den zugrunde liegenden Gegebenheiten angepasst. Nach einer Anwendung von CSM (Varimin) sind somit auch gewisse, obgleich meist nur geringfügige, Interkongruenzen der Faktoren feststellbar. Nach einer Anwendung von SSM (Varimax) findet man in der Regel deutlich höhere Interkongruenzen. Varimax erreicht somit trotz orthogonaler Rotationen das Ziel unkorrelierter Faktoren weniger zufriedenstellend als Varimin.

$$V = n\sum_{p=1}^{m}\sum_{j=1}^{n}(b_{jp}/h_j)^4 - \sum_{p=1}^{m}(\sum_{j=1}^{n}b_{jp}{}^2/h_j{}^2)^2 \qquad (1)$$

h = Kommunalität eines Faktors
b = Ladungshöhe eines Faktors

p = Laufindex für Faktoren von 1 bis m
j = Laufindex für Variablen von 1 bis n

Für das *Varimin*-Verfahren werden die Koordinaten-Drehungen mit dem Ziel einer Varianz-Verminderung vorgenommen, iterativ so lange, bis die Summe der Ladungsquadrate, das Kriterium V, sich nicht mehr vermindern lässt.

Angenommen eine initiale Struktur der Abbildung 2.03A liege vor. Die Anwendung von Varimax transformiert die Struktur 2.03A in die Struktur 2.03C. In 2.03C laufen die Koordinaten durch die Punktwolken, für die einzelnen Faktoren wird die Summe ihrer Ladungsquadrate maximiert. Die Anwendung von Varimin transformiert 2.03A in 2.03B. Der Abstand der Variablenwolken von den Koordinaten wird so weit wie möglich vergrößert. Die Summe der Ladungsquadrate wird somit für die Faktoren minimiert.

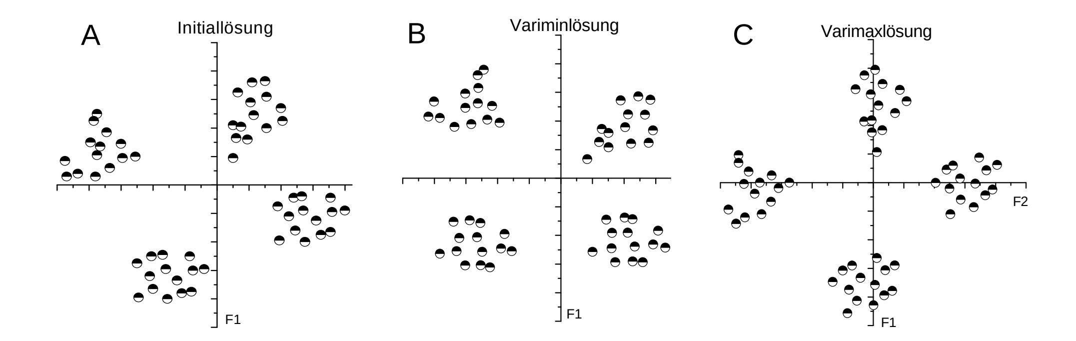


Abbildung 2.03

Zweifaktorielle Initiallösung mit fingierten Variablen (A),
die Varimin-rotiert (B) und Varimax-rotiert wurden (C).



Frage I: Wie lassen sich Varimin-transformierte Faktoren deuten?

Die SS-Rotation einer Faktorenlösung wird herkömmlicherweise deshalb für notwendig gehalten, weil die initiale Faktorenlösung komplex ist und Faktoren angeblich nicht gedeutet werden können, wenn die manifesten Variablen, auf die man sich bei der Faktorendeutung stützen muss, mehrfaktoriell geladen sind (Guilford, 1952, p. 27, Burroughs & Miller, 1961, p. 37, Überla 1971, p.175, Gorsuch, 1974, p. 162, Comrey, 1978, p.653 f., Reise et al., 2000, p. 292).

Die Behauptung, dass die faktorielle Komplexität der Variablen eine Deutung der Faktoren verhindere oder erschwere, lässt sich jedoch nicht aufrecht erhalten, wenn man das Konzept der ‚distinctive features' methodisch nutzt. Das Konzept wurde in der Phonologie von Jacobson & Halle (1956) entworfen und fand in dieser Disziplin als ein bedeutender methodischer Fortschritt bis heute weithin Anerkennung. In generalisierter Form ist seine Anwendung auf andere als phonologische und linguistische Inhalte übertragbar.

Weitere Details: Phoneme lassen sich mithilfe von ‚features' (Merkmale) unterscheiden, sofern diese eine distinktive Funktion für die untersuchte Sprache besitzen. Für jedes Phonem hat man auf dem Wege der Selbstbeobachtung und mithilfe motorischer Messungen drei Klassen von Merkmalen als funktional relevant erkannt, welche die Artikulation der Phoneme spezifizieren: die Artikulationsdauer (kurz vs. lang) [49], die Sonorität oder Stimmbeteiligung (stimmhaft, stimmlos) und den Artikulationsort (bilabial, labio-dental usw. bis uvular). Hier lässt sich die Begrifflichkeit der Faktorenanalyse anwenden, denn man kann sagen: Phoneme sind manifeste Einheiten mit jeweils drei distinktiven Merkmalen oder zugrunde liegenden (‚latenten') Varianzquellen. Jedes Phonem ist durch drei Varianzquellen zu beschreiben, zum Beispiel ist [b] ein kurz artikuliertes (plosives), ein stimmhaftes und ein bilabiales (mit den Lippen gebildetes) Phonem.

---

[49] Die zusätzlich übliche Unterscheidung hinsichtlich Artikulationsmodus (plosiv, frikativ, nasal usw.) ist spezieller und kann mit der hier gewählten Phonem-Stichprobe nicht ermittelt werden. Bei [m] und [f] handelt es sich um Phoneme mit längerer Artikulationsdauer (im Unterschied z. B. zu den kurz artikulierten plosiven Lauten [b], [p], [t],[k]), doch [m] gehört mit seiner Artikulation zu den nasalen Lauten, [f] gehört zu den frikativen Lauten. Da alle plosiven Phoneme kurz sind, hätte man auch eine bipolare Kategorie ‚plosive' vs ‚nicht-plosive' Phoneme bilden und diese als Faktor im Beurteilungsversuch erwarten können.

Zu fragen bleibt aber, wie z. B. beim Phonem [b] die drei Merkmale methodisch aufgespürt werden können. Dies geschieht durch den *Minimalpaar-Vergleich* (*minimal pair comparison*). Zur Paarbildung für [b] ist eine ähnliche andere Einheit der gleichen Domäne mit heranzuziehen. Unter den Phonemen der deutschen Sprache hat das [p] mit dem [b] zwei der oben genannten Merkmale gemeinsam (Artikulationsort und -dauer). Nur hinsichtlich der Stimmhaftigkeit unterscheiden sich diese beiden Phoneme. Wenn die Phoneme der deutschen Sprache als Variablen z. B. für eine Beurteilung ihrer Ähnlichkeiten/Verschiedenheiten verwendet werden, dann sollte eine Faktorenanalyse der Phonem-Interkorrelationen, welche Ähnlichkeitsmaße darstellen, mit anschließender Variminrotation Faktoren erkennen lassen, die den bekannten drei distinktiven Merkmalkategorien entsprechen. Die Phoneme [b] und [p] könnten z. B. bei den Faktoren F_1 und F_2 gleiche Ladungen aufweisen, bei F_3 aber müssten dann die Ladungen markant auseinander treten. Die Aufgabe des Faktorenanalytikers bestünde lediglich darin, den 'minimalen' Unterschied zwischen [b] und [p] für F_3 zu deuten: F_3 wäre dann als Effekt der Stimmhaftigkeit/Stimmlosigkeit anzusehen. Die Deutung würde sich sichern lassen, wenn weitere minimale Paare aus dem Datensatz gebildet werden, z. B. die Paare [d] vs. [t] sowie [g] vs. [k], bei denen der in gleicher Weise zu deutende minimale Unterschied für F_3 auftritt wie bei [b] und [p], während für F_1 und F_2 bei den Paaren [d] vs. [t] sowie [g] vs. [k] wieder unterschiedslose Ausprägungen vorliegen, wobei für diese Phonempaare bei beliebigen anderen Merkmalen wieder gleiche F_1- und F_2 -Ausprägungen vorliegen dürfen (nicht müssen) als beim Vergleich von [b] und [p]. [50]

Analyse 1: Beurteilung von Phonem-Ähnlichkeiten

Diese Überlegungen wurden für ein Experiment verwertet, an dem zwei deutschsprachige Studierende teilnahmen, eine Psychologiestudentin mit offensichtlicher Sprachbegabung und eine ebenfalls deutsche Linguistikstudentin höheren Semesters, zu deren Ausbildung aber die Phonologie gehörte. Die Probandinnen hatten 10 Phomene der deutschen Sprache ([b], [d], [f], [g], [k], [m], [n], [p], [t], [v] auf bipolaren Likert-Skalen mit sieben Stufen (Mittelstufe = 0) nach Ähnlichkeit zu

[50] Die hier bevorzugte faktorenanalytische Forschungsstrategie lässt sich unmittelbar verknüpfen mit theoretischen und empirischen Ansätzen, in denen Begriffe und andere kognitiv repräsentierte Objekte als Bündel oder Strukturen mehr oder weniger latenter Komponenten aufzufassen sind ('Komponentenmodell' von Einstellungsobjekten, Feger, 1971, 'Feature Pattern Analysis', Feger & Brehm, 2001). Auch läuft sie in die gleiche Richtung wie die Bemühungen um das Phänomen des Ähnlichkeitserlebens, zu dessen Erklärung bei Shepard, 1974, ‚hidden structures' und bei Tversky, 1977, ‚collections of features' analysiert wurden.

beurteilen ([v] wird ausgesprochen wie [w] in z. B. ‚wo'). Als Skalenpole wurden die zu beurteilenden 10 Phoneme paarweise in allen Kombinationen (= 45 Kombinationen) verwendet, die zufällig aufeinander folgten, z. B. waren die Phoneme auf den Skalen
[d] 3---2---1---0---1---2---3 [m]
[b] 3---2---1---0---1---2---3 [d]
und weiteren 43 Skalen nach Ähnlichkeit einzustufen. Wenn das zu beurteilende Phonem gleichzeitig auch einen Skalenpol definiert, läuft das auf ein Identitätsurteil hinaus, z. B. ist [d] auf der Skala
[d] 3—2—1—0—1—2—3 [m]
dem linken Skalenpol maximal ‚ähnlich' (mit ihm identisch). In solchen Fällen war laut Instruktion die Stufe 3 (maximale Ähnlichkeit) anzukreuzen.

Die zehn Urteilsprofile mit je 45 Urteilen wurden sodann für jede Probandin interkorreliert und einer Principal Component-Analyse (PCA) unterzogen. Die ersten drei extrahierten Faktoren (= Komponenten) wurden für die Varimin-Rotation berücksichtigt, da drei interpretierbare Faktoren erwartet wurden (Eigenwerte der ersten fünf Faktoren: 2.23, 1.58, 1.40, 1.12, 1.05 für Vp 1 und 4.86, 1.68, 1.09, 1.03, 0.59 für Vp2).

Die Güte der Varimin-Rotation wird danach bemessen, wie gut die Faktorladungen mit den zu erwartenden Klassifizierungen übereinstimmen. Nach der phonologischen Klassifizierung haben die folgenden Phoneme eine kurze Artikulationsdauer: [b], [d], [g], [k], [p], [t] (man nennt sie ‚plosiv'), die Phoneme [f], [m], [n], [v] werden mit anhaltendem Luftstrom, also länger artikuliert). Die Phoneme [b], [d], [g], [m], [n], [v] sind stimmhaft, stimmlos sind [f], [k],[p], [t]. Während sich nach der Artikulationsdauer und Stimmbeteiligung eine bipolare Unterteilung ohne zusätzliche Differenzierung ergibt, liegen beim Artikulationsort vier Stufen vor. Von der Artikulation ‚weit vorn' bis ‚weit hinten' sind aufzuführen (1) die bilabialen Phoneme [b], [m], [p], (2) die labiodentalen Phoneme [f], [v], (3) die alveolaren [t], [d], [n], und (4) die velaren [k], [g]. Die Ladungen der aus den Urteilsdaten der Probandinnen ermittelten Varimin-rotierten Faktoren werden sodann punktbiserial korreliert mit den linguistischen Variablen Artikulationsdauer (kurz = 1 und lang = 2) und Stimmhaftigkeit (stimmlos = 1 und stimmhaft = 2). Mit den vier ordinalen Stufen des Artikulationsortes (ganz vorn, 1, bis ganz hinten, 4) werden die Faktorladungen Produktmoment-korreliert.

Die Korrelationen zeigt Tabelle 2.02. Man sieht, dass mit Ausnahme der Korrelation für den Artikulationsort bei der Psychologie-Probandin (r = .671) die Korrelationen sehr hoch sind, sie liegen über .90. F_1 repräsentiert bei beiden Probandinnen die Artikulationsdauer, F_2 repräsentiert die Stimmhaftigkeit. Offenbar hat die Linguistik-Studentin aufgrund ihrer Ausbildung ein feineres Empfinden erworben, was die Wahrnehmung des Artikulationsortes im Mund betrifft (die Tucker-

Kongruenz für F_3 bei den beiden Versuchsteilnehmerinnen beträgt nur .574. Da aber die Ladungen bei F_3 für die Linguistik-Studentin wesentlich höher ausfallen als bei der Psychologie-Studentin, ergibt sich nach einer Mittelung der beiden Datensätze mithilfe von Fishers Z-Transformation auch für F_3 eine Korrelation mit dem Artikulationsort von $r = .922$, so dass weiterhin die Ergebnisse der vereinten beiden Datensätze für alle drei Faktoren dargestellt werden dürfen.

Tabelle 2.02

Korrelationen der Faktorladungen mit objektiven Rangwerten
sowie die Tucker-Kongruenz der Faktoren von zwei Probandinnen.

	Faktoren	Korrelationen der Faktorladungen mit objektiven Rangwerten			Kongruenz der Faktoren der Vpn
		Vp Psych	Vp Lingui	beide Vpn	
F1	kurz-lang	.972	.990	.996	.943
F2	stimmhaft-stimmlos	.917	.957	.966	.936
F3	Artikulations-ort	.671	.976	.922	.574

Abbildung 2.04 zeigt die Varimin-Ergebnisse. Positive Faktorladungen werden durch dunkle Kreise, negative durch hellere Kreise dargestellt, die variablen Kreisflächen entsprechen den variablen Ladungshöhen. Eine Null-Ladung entspräche einem ausdehnungslosen Punkt. Wäre die Deutung der drei Faktoren
nicht schon durch die Korrelationen mit Expertenurteilen erfolgt, hätte man für F_1 die minimalen Paare [t] vs. [f] und [d] vs. [n] bilden können, wodurch die Artikulationsdauer als distinktives Merkmal in Erscheinung tritt. Bei F_2 wären die minimalen Paare ‚Stimmbeteiligung ja' vs. ‚Stimmbeteiligung nein' aufgefallen: [b] vs. [p], [d] vs. [t], [g] vs. [k] und [v] vs. [f]. Bei F_3 hätten die minimalen Paare [p] vs. [k] und [b] vs. [g] den vorn-hinten Kontrast des Artikulationsortes manifestiert.

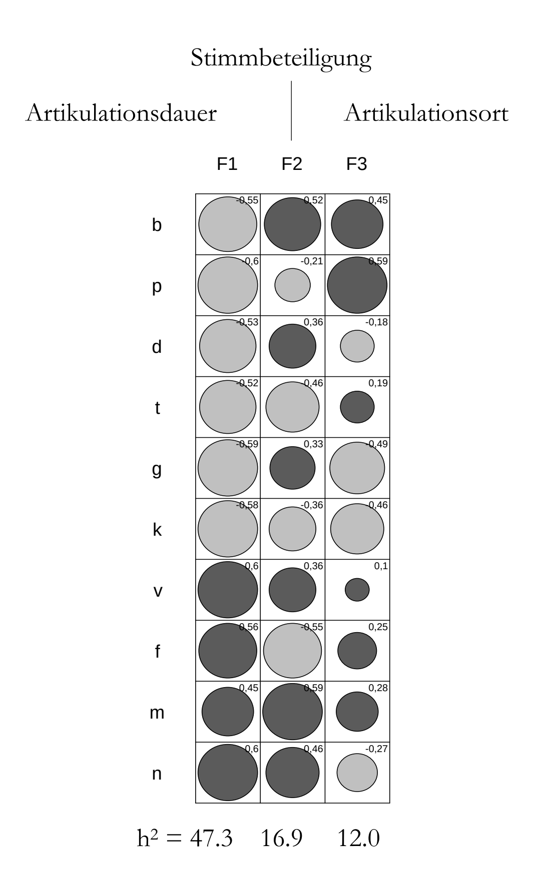


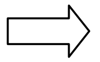
Abbildung 2.04

Varimin-transformierte Faktorladungen von zehn deutschen Phonemen (aufgrund von Ähnlichkeitsbeurteilungen).

Das Ergebnis der Varimax-Rotation der Faktoren wird ohne Abbildung mitgeteilt, das der Varimin-Rotation nur von der Linguistik-Probandin, weil eine Deutung bei ihr weniger klar ist als bei der Psychologie-Probandin. Durch Varimax-F_1 werden bei der Linguistk-Probandin die *langen stimmhaften* Phoneme [m], [n] und [v], die hier ein positives Vorzeichen haben, geclustert, mit negativem Vorzeichen gruppieren sich bei F_1 die *kurzen stimmlosen* Phoneme [t] und [k]. F_2 gruppiert mit positivem Vorzeichen die *kurzen stimmhaften* Phoneme [g] und [d]. Eine negative F_2-Ladung hat das *lange stimmlose* [f]. Der unipolare Faktor F_3 gruppiert die bilabialen plosiven Phoneme [p] und [b]. Es ist offenkundig, dass die Varimax-Rotation

zwar einander ähnliche Phoneme clustert (die mit Plus-Ladung) und auch die jeweils gegensätzlichen Cluster zusammen bringt (die mit Minus-Ladung). Doch an solchen Clustern, zu denen immer mehrere phonetische Merkmale beitragen, kann man kaum interessiert sein. Denn gesucht werden die latenten Quellen der Kovarianz. Bei der Herstellung von SS-Strukturen, also Clustern, werden die effektiven Varianzquellen *nicht* aufgefunden, sie werden verschleiert.

Eine Deutung von Varimin-transformierten Faktoren ist demnach unproblematisch, sofern minimale Paare vorliegen. Die Deutung von Faktoren der SSM-orientierten Faktorenanalysen ist in der Regel entweder unmöglich oder ungleich schwieriger, weil bei der Deutung dieser Faktoren *globale Ähnlichkeiten* unter den faktoriell geclusterten Variablen beschrieben werden müssen, wobei immer mehrere zugrunde liegende Merkmale zur Ähnlichkeit bzw. Verschiedenheit beitragen. Bei der Deutung von Faktoren, die durch Complex Structure Modellierung (CSM) entstehen, sind nur Unterschiede zwischen jeweils zwei gepaarten Variablen bezüglich eines einzigen Merkmals zu ermitteln. Einfacher also ist das Ergebnis einer CSM als das einer SSM – scheinbar paradoxerweise.



Frage II: Wird die Komplexität der Kovarianzquellen nicht schon durch die Initiallösung hinreichend repräsentiert?

Man kann bezweifeln, dass es nötig sein soll, die in der Initiallösung schon vorliegende Komplexität durch eine Varimin-Rotation noch zu vergrößern. Das dafür vorgebrachte Argument könnte nicht ausreichen (das Argument war: realitätsverzerrende Bevorteilung der zuerst extrahierten Faktoren gegenüber später extrahierten). Die bislang vorgebrachte Kritik könne das Konzept der Einfachstruktur allenfalls nur relativieren. Die Forscher dürften durch die vorliegende Kritik auf den vernachlässigten Informationswert der Initialstruktur aufmerksam gemacht werden, nicht jedoch würde eine Varimin-Rotation der Initialfaktoren erforderlich sein, die die Komplexität der initialen Lösung erhöht.

Diesen Einwand wird man wohl nur mit geeeigneten empirischen Prüfergebnisseen entkräften können. Die Ergebnisse Varimin-rotierter Lösungen, die sich von den Ergebnissen initialer Lösungen unterscheiden, müssten sich als durchweg leichter interpretierbar erweisen, d.h. als besser integrierbar in den jeweils vorliegenden Wissenskontext im Vergleich zu initialen Lösungen, nicht also nur im Vergleich zu Varimax-rotierten Lösungen. Die relative Modelliergüte der Lösungen wäre für die anstehende Frage letztendlich entscheidend. Unter zahlreichen empirischen Belegen, bei denen die Modelliergüte einer Varimin-Rotation die der Initiallösung übertrifft, werden zunächst die Ergebnisse einer Analyse der Merkmale beurteilter Münzen ausgewählt.

Analyse 2: Ähnlichkeitsbeurteilung britischer Münzen

Britische Münzen als Urteilsobjekte eignen sich als manifeste Variablen einer faktoriellen Analyse insbesondere deshalb, weil sie eindeutige Merkmale (Varianzquellen ihres phänomenalen Eindrucks) besitzen, nämlich Größe, Form und Farbe. Diese stützen sich auf physikalische Reizeigenschaften, ihre Feststellung ist nicht auf Eindrucksurteile ohne Reizgegebenheiten angewiesen (s. Abbildung 2.05). Die nach ihrem Geldwert benachbarten Münzen sind größenverschieden (z.B. *1 Pence* klein, *2 Pence* groß, *5 Pence* klein, *10 Pence* groß usw.). Dadurch ergeben sich natürliche Paare, die sich wiederum von anderen Paaren entweder durch die Metallfarbe (*5* und *10 Pence* z.B. sind silbern, *1* und *2 Pence* sind nicht silbern) oder durch ihre Form voneinander unterscheiden (z.B. sind *5* und *10 Pence* rund, *20* und *50 Pence* sind siebeneckig). Diese Merkmale liegen der Kovarianz der Ähnlichkeitsurteile zugrunde, die beim Vergleich der untersuchten Objekte entstehen. Die Durchmesser und Gewichtsmaße der Münzen sind aus Tabelle 2.02 ersichtlich.

Farbe	**Pence**	Größe	Form
nicht silbern	1	kleiner	*rund*
nicht silbern	2	größer	*rund*
silbern	5	kleiner	*rund*
silbern	10	größer	*rund*
silbern	20	kleiner	**nicht rund**
silbern	50	größer	**nicht rund**
	Pound		
nicht silbern	1	kleiner	*rund*
nicht silbern	2	größer	*rund*

Abbildung 2.05
Das System der britischen Münzen
mit den Eigenschaften Farbe, Größe und Form

Tabelle 2.02

Durchmesser und Gewicht der Münzen

	1 Penny	2 Pence	5 Pence	10 Pence	20 Pence	50 Pence	1 Pound	2 Pound
Gewicht (g)	3.28	7.10	3.52	6.50	5.00	8.20	9.48	11.92
Durchmesser (mm)	20	25	18	24	20	28	22	28

Das Experiment: Eine in Cambridge studierende deutsche Studentin wurde gebeten, die acht Münzen der gegenwärtigen britischen Währung nach Ähnlichkeit einzustufen (*1, 2, 5, 10, 20, 50 Pence* und *1* und *2 Pound*). Die Münzen wurden auf Karton aufgeklebt und in Paaren dargeboten: Jede Münze kam mit jeder anderen Münze gepaart vor, angefangen mit den Paaren *1 Penny* [51] *– 2 Pence*, *1 Pence – 5 Pence* usw. bis zum Paar *1 Pound – 2 Pound*, insgesamt 28 Paare.

Die Studentin hatte jede der acht Münzen in die Hand zu nehmen und mit den beiden aufgeklebten Münzen der Karton-Vorlage zu vergleichen. Auf einer bipolaren Siebenpunkt-Skala hatte sie anzugeben, ob die Münze in ihrer Hand eher der links oder der rechts aufgeklebten Münze ähnlich war. Zum Beispiel könnte ihr die 50 Pence-Münze eher der *5-Pence* Münze als der *1 Pound*-Münze ähnlich erscheinen, worauf sie dann auf der Skala *5 Pence – 1 Pound* einen Skalenpunkt in der Nähe des *5 Pence*-Pols ankreuzen würde. Ein ganzheitliches Urteil wurde erbeten, alle Ähnlichkeits- und Verschiedenheitsgründe sollten gleichgewichtig beachtet wer den. Vom finanziellen Münzenwert sollte abgesehen werden.

Auf diese Weise kam für jede Münze ein Urteilsprofil mit 28 Einzelurteilen zustande. Die Urteilsprofile der acht Münzen wurden interkorreliert, die Korrelationsmatrix wurde einer PCA unterzogen, die extrahierten Faktoren anschließend der Varimin- und Varimax-Rotation zugeführt. Die Erwartung war, dass sich das Merkmalsystem der britischen Münzen nach einer Varimin-Rotation, nicht aber nach einer Varimax-Rotation, faktoriell durchsetzen würde, und dass das Varimin-rotierte Ergebnis Münzen-Merkmale besser hervortreten lassen würde als die initiale Lösung.

[51] Umgangsprachlich ersetzt oft *one pence* das eigentlich richtige *one penny*.

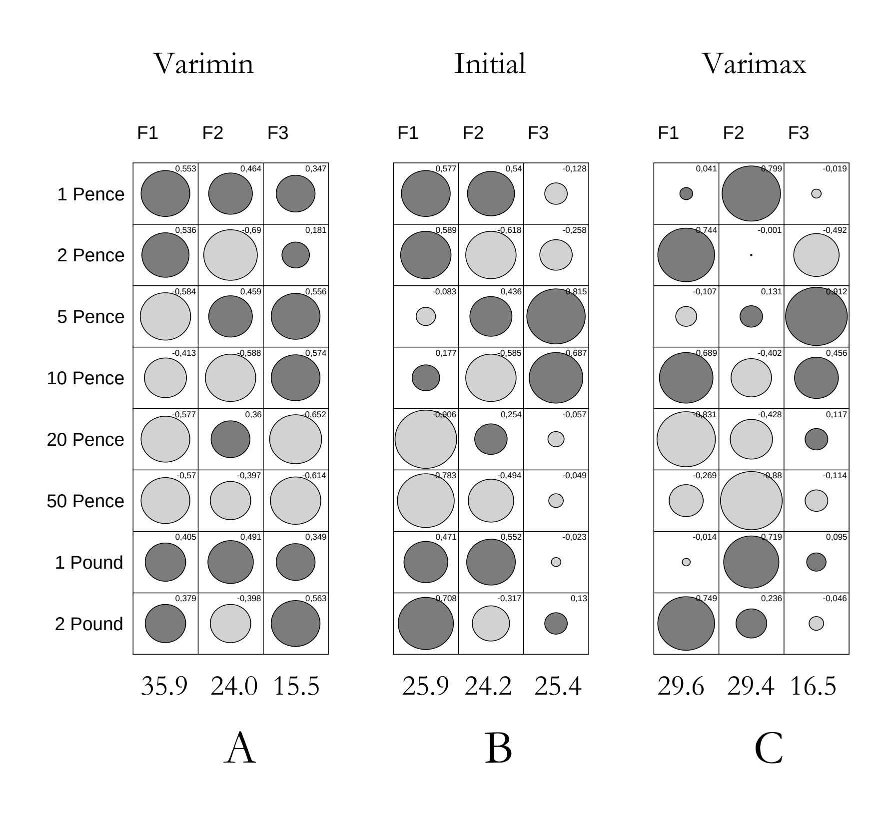


Abbildung 2.06

Varimin- (A), intiale (B) und Varimax-transformierte Lösung (C)
mit Ähnlichkeitsdaten britischer Münzen

Das Ergebnis (Abbildung 2.06) fällt für Varimin erwartungsentsprechend aus: [52] Die Münzen von silberner Farbe heben sich auf dem ersten bipolaren Faktor von den Münzen mit nicht-silberner Farbe eindeutig ab. Dies drücken die Ladungsvorzeichen aus, die Ladungshöhen unterscheiden sich nur wenig. Dasselbe gilt für

[52] Der Durchmesser der Kreise entspricht der Ladungshöhe, die die Kreise veranschaulichen, die hellen Kreise repräsentieren positive Ladungen, die dunklen negative Ladungen. Numerische Werte befinden sich klein gedruckt in den Zellen der Matrix. Unterhalb der Spalten befinden sich die Kommunalitätsprozente der Faktoren.

den zweiten Varimin-Faktor, der offensichtlich die Größen-Unterschiede der Münzen wiedergibt. Im dritten Faktor schlägt sich offenbar der Varianzanteil der Münzenform nieder. [53] Eine Deutung mithilfe minutiöser Minimalpaar-Vergleiche kann man sich bei der
hier vorliegenden Transparenz der ‚latenten' Bedingungslage ersparen.

Zur Initiallösung: Unter den Faktoren der Initiallösung lässt sich F_2 als Manifestation der Münzengröße deuten. In F_1 könnte man den Farbfaktor vermuten, doch lassen sich die Ladungen von 5 und 10 Pence nicht mit der Münzenfarbe in Zusammenhang bringen. Eine Deutung von F_3 als Formmerkmal ist ausgeschlossen, F_3 der Initiallösung bleibt rätselhaft. Das hier deutlich gewordene defizitäre Ergebnis der Initiallösung führt mit vielen anderen nicht berichteten Ergebnissen aus anderen Untersuchungen zu der praktischen Schlussfolgerung, dass PCA-Faktoren grundsätzlich nach Varimin rotiert werden sollten, auch dann, wenn die Varimin-rotierte Lösung von der initialen nicht sehr abweichen sollte – was bei Faktor-Lösungen mit mehr als zwei Faktoren ohnehin selten vorkommt.

Zur Varimax-Lösung: Das Ideal einer einfaktoriellen Ladung wird nur etwa bei der Hälfte der Münzen erreicht. Die 10 Pence und 20 Pence-Münzen weichen vom solitären Ladungsideal erheblich ab. Multiple Faktorladungen sind in der SS-Praxis seit jeher bekannt, sie wurden immer als ein leidiges Problem hingenommen. Schwerer wiegt die Tatsache, dass Varimax eine deskriptive Einfachheit auf der inhaltlichen Ebene nicht erreicht. Weder Farbe, noch Form, noch Größe werden durch die Varimax-Faktoren repräsentiert, nicht einmal andeutungsweise.

Um die Generalisierbarkeit des Ergebnisses zu prüfen, wurde das Experiment wiederholt mit acht Studierenden, die mit britischen Münzen keinen Umgang hatten. Die über die Versuchsteilnehmer gemittelten Urteilswerte wurden genau so analysiert wie die individuellen Werte der Cambridge-Studentin (T. S.). Es zeigte sich, dass mit dem Kriterium des Eigenwertverlaufs 2.07, 1.90. 1.30, 1.20 vier Faktoren extrahierbar waren, bei T. S. waren es nur drei Faktoren (Eigenwerte: 2.87, 1.92, 1.24, 0.97). Die faktoriellen Kongruenzen der zu vergleichenden Lösungen, von T. S. und Studenten, waren für F_1 (Farbfaktor) und F_2 (Größenfaktor) [54] beachtlich, sie betrugen .981 bzw. .973. Doch fand sich für F_3 von T.S.

---

[53] Die leicht variierenden Ladungshöhen bei F_2 (am meisten fällt das *2 Pence*-Stück aus der Reihe) könnten mit münzenspezifischen subjektiven Gewichtungen der Merkmale zusammenhängen. Tatsächlich wirkt das *2-Pence* Stück angesichts seines geringen Wertes übermäßig groß, so dass bei ihm im Vergleich mit anderen Münzen wahrscheinlich der Größeneindruck stärker als der Formeindruck zur Geltung kommt.

[54] Die Interpretation von Varimin-F_2 als Faktor der anschaulichen Münzengröße wird gestützt durch Korrelationen der F_2-Ladungen mit der gemessenen

weder bei F_3 noch bei F_4 der Studenten eine Spur von Kongruenz (.312 bzw. .224). Es stellte sich heraus, dass bei den Studenten die acht F_3 - Gewichte mit den acht Rangplätzen des Geldwertes der Münzen hoch korrelierten (.88), während für T. S. die höchste Korrelation mit dem Geldwert nur .14 betrug (bei F_1). Offensichtlich haben sich die Studenten bei ihren Ähnlichkeitsurteilen außer von Farbe und Größe noch vom Geldwert der Münzen beeinflussen lassen, während die Münzenform , die sich bei T. S. als F_3 manifestiert hatte, von ihnen unbeachtet blieb. Dass bei den Ähnlichkeitsurteilen der Geldwert der Münzen nicht berücksichtigt werden sollte, war zwar in der schriftlichen Instruktion für die Studenten gesagt, aber nicht wie in der mündlichen Instruktion für T. S. besonders betont worden.

⇨ Frage III: Lassen sich faktorielle SS-Lösungen nicht doch oft befriedigend interpretieren, wie könnte man sie sonst ständig verwenden?

Die Variablen der bisher berichteten Untersuchungen waren Objekte, die die Probanden aufgrund wahrgenommener Merkmale nach Ähnlichkeit zu beurteilen hatten. Daten dieser Art wurden selten faktorenanalytisch verarbeitet, denn mit der SS-orientierten Methode, die solitäre Ladungsverhältnisse unter den Variablen zum Ideal erhebt, gewinnt man mit den Ähnlichkeitswerten, die man bei Objektvergleichen erhält, von vorne herein für SSM ungeeignete, d. h. komplexe (mehrdimensionale) Objektbeschreibungen. So weicht man auf alternative Verfahren aus, z. B. auf MDS (multidimensionale Skalierung).

Ein Haupt-Einsatzgebiet für Faktorenanalysen in der Psychologie sind Urteilsdaten über Menschen (Selbst- und Fremdbeurteilungen), Urteile über Persönlichkeitseigenschaften, Verhaltensdispositionen, Einstellungen usw.. Die Varianz der dabei verwendeten Variablen, d. h. der für solche Erhebungen verwendeten verbalen Items, ergibt sich durch Messwiederholungen über urteilende Personen. Faktorenanalysen mit solchen Daten führen, wenn sie SS-orientiert vorgehen, zur faktoriellen Gruppierung verbal-semantischer Urteilseinheiten. Bei diesen kann man sich zur Einschätzung der Modelliergüte nicht wie bei Objektbeurteilungen auf anschauliche Kriterien stützen. Semantische Urteilseinheiten aus der Erlebnis- und Verhaltensdomäne sind weniger streng voneinander geschieden, sie lassen sich immer irgendwie deuten, die faktoriell gruppierten Variablen machen einen mehr oder weniger ‚sinnvollen' Eindruck. Das Ungenügen der SSM wird bei solchen Variablen nicht erkannt. Zwar wird mitunter von signifikanten Korrelationen

---

Münzengröße: r = .81 (T. S.), r = .85 (Studenten). Das Gewicht der Münzen in Gramm korreliert mit den Faktorladungen nur geringfügig, obwohl es mit ihrer Größe in cm bei r = .80 korreliert.

zwischen Faktorscores der Selbstbeurteilungen und objektiven Verhaltensdaten berichtet, doch wird dem Faktorenanalytiker die konzeptuelle Arbeit, die er beim Ermitteln latenter Bedingungen der Faktorenextraktion aufbringen muss, auch bei anschließender Verwendung von Außenkriterien nicht erspart.

In der folgenden Untersuchung werden Korrelationen von Variablen faktorisiert, bei denen Messwiederholungen über urteilende Personen eine Rolle spielen. Auf Probleme semantischer Deutung stoßen wir bei diesen Daten noch nicht, denn die Variablen sind inhaltsarme Adverbien des Ausprägungsgrades (Formwörter). Sie stammen aus einer Untersuchung von Carl (1968) über das Antwortstil-Verhalten. Mit den an diesen Daten gewonnenen Ergebnissen lässt sich aufzeigen, wie SS-Faktoren und ihre Deutungen entstehen. Zu diesem Zweck werden die mit denselben Daten ermittelten Variminfaktoren herangezogen.

Analyse 3: Differenzierung von Antwortstilen bei der Beantwortung von Fragebögen (Daten von Carl).

Das Ziel Carls (1968) war die Ermittlung von Antwort-Tendenzen (response sets) bei der Verwendung von Stufenskalen. Er hatte von 100 Personen Fünf-Punkt-Likert-Einstufungen für 580 Items aus dem *Minnesota Multiphasic Personality Inventory* (MMPI) erhoben und für jede Person über alle Items, ohne Rücksicht auf deren Inhalt zu nehmen, die Anzahl der fünf Einstufungen summiert, die von ‚stark zustimmend' bis ‚stark ablehnend' reichten (die Zwischenstufen der Skalen waren nicht verbalisiert worden). Carl hatte ausgeschlossen, dass aufgrund der gegenseitigen Abhängigkeit der Stufen-Urteile Artefakt-Korrelationen vorkamen: Er ermittelte die Verwendungshäufigkeiten für die fünf Antwortstufen mit fünf aus dem Gesamtpool gewonnenen Teilstichproben von Items, die inhaltlich parallelisiert worden waren. Aus der Interkorrelationsmatrix Carls, die auch Variablen aus Skalen mit mehr Stufen enthielt (er hatte z.B. auch Sieben-Punkt-Skalen verwendet), wurde für den vorliegenden Zweck die Matrix für das Format mit fünf Urteilsstufen separiert. Für diese wurde mithilfe einer PCA die initiale Faktorenstruktur ermittelt, die dann mit Varimin und zum Vergleich mit Varimax rotiert wurde.

Bei dieser Untersuchung sind mit fachlichem Vorwissen bestimmte Ergebnisse erwartbar. Denn bekannt ist die *Ja-Sage-Tendenz* (Akquieszenz), die einen Faktor hervorbringen müsste, bei dem die beiden zustimmenden Antwortstufen positiv und die beiden ablehnenden Antwortstufen negativ geladen sind. Bekannt ist auch die *Extremantwort-Tendenz*, die einen Faktor hervorbringen müsste, bei dem die beiden Extrem-Stufen der Zustimmung und Ablehnung positiv und die dazwischen liegenden Stufen negativ geladen sein müssten (evtl. nach Vorzeichenum-

kehr, die Polungsrichtung ist irrelevant). Mindestens diese beiden Faktoren also müsste eine Nachbehandlung der initialen Faktoren durch Varimin zutage fördern.

Das *Varimin-Ergebnis* zeigt Abbildung 2.07B. Man erkennt in F_3 den erwarteten Akquieszenz-Faktor und in F_1 den Extremantwort-Faktor. Allerdings ist ein weiterer substanzieller Faktor F_2 vorhanden, der nicht erwartet wurde und der zur Deutung herausfordert. Man kann hier bei Herrmann (1965) anknüpfen, der die von ihm so genannte *Urteilsnuanciertheit* als Variante eines Antwortstils bei Fragebogen-Untersuchungen identifizierte. Demnach unterscheiden sich Personen bei Stufenskalen auch darin, inwieweit sie über eine zustimmende vs. ablehnende Entscheidung hinausgehend zwischen ‚Zustimmung' und ‚starker Zustimmung' sowie zwischen ‚Ablehnung' und ‚starker Ablehnung' und auch zwischen ‚Zustimmung/Ablehnung' einerseits und ‚Unentschieden' bzw. ‚Weiß nicht' andererseits differenzieren. Diese Antwortstil-Variante wird in der Literatur seltener diskutiert (gelegentlich unter der Bezeichnung ‚Urteilsvariabilität'„response variance'), vermutlich weil durch sie die Ergebnisse weniger verzerrt werden als durch andere Antwort-Tendenzen (Hinz et al., 2003). Der zur Klärung anstehende Faktor F_2 der Varimin-Analyse lässt sich m. E. risikolos als Ausdruck der ‚Urteilsnuanciertheit' deuten.

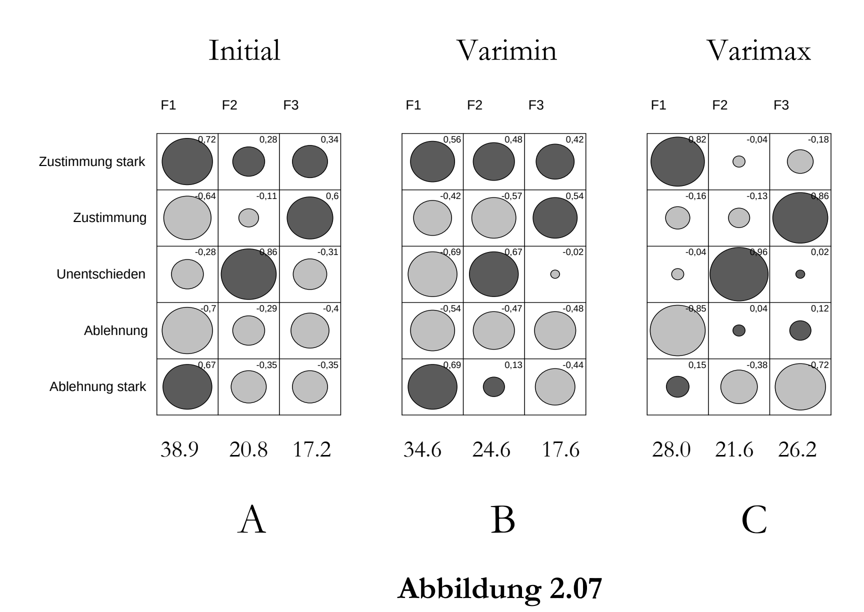


Abbildung 2.07

Initiale (A), Varimin- (B) und Varimaxstruktur (C)
der Response Set-Daten von Carl

Sind die drei Response Sets schon in der *initialen* Faktorenstruktur zu finden (s. Grafik 7A)? Die initiale Faktorstruktur ist der Varimin-Faktorstruktur sehr ähnlich. Faktor F_1 ist in den beiden Lösungen nahezu identisch, also ist auch der *initiale* F_1-Faktor als Ausdruck der Extremantwort-Tendenz zu deuten. Doch bei F_3, Akquieszenz, ist die initiale Lösung weniger gut, da hier die mittlere Kategorie der Urteils-Enthaltung eine beachtliche negative Ladung aufweist, während doch einer Urteils-Enthaltung andere Bedingungen zugrunde liegen sollten als einem Urteil, das eine Ablehnung beinhaltet. Auch für F_2 ist die Varimin-Lösung prägnanter. Bei der initialen Lösung wird bei F_2 zwischen ‚Ablehnung' und ‚starke Ablehnung' nicht differenziert. Auch ist die numerische Differenz zwischen ‚Zustimmng ' und ‚starke Zustimmung' in der initialen Lösung deutlich schwächer als bei der Varimin-Lösung. Von Urteilsnuanciertheit kann man beim initialen F_2 -Faktor kaum sprechen.

Unser Hauptinteresse aber gilt hier dem *Varimax*-Ergebnis (Grafik 7C). Wie lässt es sich interpretieren? Darüber gibt ein Vergleich der Varimax- mit der Varimin-Lösung (vgl. Abbildung 2.07C mit 2.07B) Aufschluss. Die Varimax-Rotation hat zur bipolaren Clusterung der Variablen ‚starke Zustimmung' (positive Ladung bei F_1) und ‚Ablehnung' geführt (negative Ladung bei F_1). Warum? Es fällt auf, dass die Vorzeichen der Ladungen von ‚starke Zustimmung' und ‚Ablehnung' bei den drei Merkmalen, die durch Varimin differenziert wurden, kontrastieren. Auch beim Varimax-Faktor F_3, der ebenfalls bipolar ist, finden sich bei den Varimin-Merkmalprofilen Kontraste, diesmal bei den Variablen ‚Zustimmung' und ‚starke Ablehnung'. Für die Urteilskategorie ‚unentschieden' gibt es bei den anderen Urteilsstufen im Varimin-Ergebnis keinen polaren Gegensatz, so bleibt ‚unentschie den' bei Varimax-F_2 ziemlich isoliert.

Diese Ergebnisse entsprechen dem, was der Varimax-Varimin Faktorenvergleich bei den Münzen erbracht hat. Kurz: Varimax-Transformationen clustern Variablen mit gleichartigen Merkmalprofilen. [55] Im Falle bipolarer Faktoren clustern sie zusätzlich, mit dem jeweils anderen Vorzeichen, die Variablen mit gegensätzlich ausgeprägten Profilen. Die Merkmalsgrundlage selbst wird nicht analysiert, eine

---

[55] Zimmermann (1953) bemerkte, dass Variablen, die in der unrotierten Lösung mit ähnlichem Profil mehrfaktoriell geladen sind, sich durch die SS-Rotation in einem Faktor vereinigen: …*a test which actually contains variance on two or more factors may appear with all of that variance confined on a single factor.*" Diesen nennt er ‚composite factor'. Der Autor hält also an der inhaltlichen Bedeutung der initialen Faktoren (Centroidfaktoren) fest, die er aber im rotierten Faktor nur noch als Faktorenbündel (‚composed') vereint sieht. *„It is my feeling that the failure to give composite factors the attention they merit must be considered either a serious oversight or a serious error or omission*" (Zimmermann, 1953, 389).

Differenzierung auf der Merkmalebene wird durch eine Varimax-Clusterung systematisch verunmöglicht. [56]

Was würde ein konventioneller Faktorenanalytiker nach Durchführung einer Varimax-Analyse der Response-Set-Daten möglicherweise publizieren? Er könnte behaupten, Akquieszenz sei nicht, wie man bislang dachte, ein einfaktorielles Konstrukt, man habe vielmehr zwischen einer Akquieszenz I mit Extremheitstendenz (F_1) und einer Akquieszenz II ohne Extremheitstendenz (F_3) zu unterscheiden. Dann könnte er noch F_2 als ‚Unentschiedenheit' interpretieren und sich darüber hinweg setzen, dass F_2 auch bei ‚starke Ablehnung' eine beachtliche negative Ladung aufweist. Er könnte der statistischen Datenverarbeitung vertrauen und glauben, drei neue psychologische Konstrukte entdeckt zu haben (Akquieszenz I und II und Unentschiedenheit). Da diese nicht sinnlos sind, würde man nicht bemerken, dass sie aber ziemlich nutzlose Variablengruppierungen darstellen, wie sie durch Anwendung des SSM fortwährend hervorgebracht werden.

Mit der folgenden Reanalyse *verbaler* Daten aus einer MDS-Untersuchung wird die unverbindliche Deutungspraxis, die bei Verwendung semantischer Einheiten nach einer Varimax-Faktorisierung nahe liegt, noch deutlicher.

Analyse 4: Semantische Merkmale bei Verwandtschaftsbezeichnungen
(Daten von Marx & Hejj).

Für 16 Verwandtschaftsbezeichnungen, die den Probanden auf Wortkarten dargeboten und von ihnen hierarchisch sortiert worden waren, ermittelten Marx & Hejj (1989) eine Matrix der Ähnlichkeiten (s. Tabelle 2.03, bei den Autoren Tabelle 2.5, S. 112). Die Autoren selbst verwendeten ein NMDS-Verfahren, um aus den Häufigkeiten der zusammen gelegten Wortkarten die semantischen Merkmale der Verwandtschaft zu gewinnen, was ihnen auch einigermaßen gelang.[57] Für den

[56] Dass die SS-Rotation nicht zur Analyse zugrunde liegender Dimensionen, sondern nur zur Clusterung ähnlicher Testvariablen führt, wurde explizit von Overall nicht nur behauptet, sondern anhand eines Analysebeispiels demonstriert: *"Rotation to simple structure can be understood as an elaborate approach to cluster analysis. It identifies clusters of tests which measure the same things, but there is no assurance that these 'same things' are simple and primary dimensions"* (p. 271). „*… there is no need to assume that simple structure factors will correspond to any particular set of fundamental dimensions of the objects* … (p. 276).

[57] Die Probanden hatten 16 Verwandtschaftswörter nach selbst gewählten Kriterien zunächst in zwei beliebig große Teilgruppen zu teilen, dann jede der bei

vorliegenden Zweck wurde die Originalmatrix der Sortierhäufigkeiten, welche als Indikatoren der Distanz oder Unähnlichkeit (UÄ) unter den 16 Begriffen zu betrachten sind, diagonal gespiegelt und zu einer Vierecksmatrix aufgefüllt. Sodann wurden die UÄ-Maße durch Ä = 1 - UÄ/1000 in Ähnlichkeitsmaße transformiert. [58] In die Diagonale wurde der aus der jeweils zugehörigen Spalte/Zeile zu entnehmende höchste Ähnlichkeitswert eingesetzt. Die Spalten der Ähnlichkeitsmatrix wurden sodann untereinander korreliert und die Interkorrelationsmatrix einer PCA unterworfen. Fünf substantielle Faktoren wurden extrahiert [59] und nach Varimin und Varimax transformiert.

Tabelle 2.03

Daten des hierarchischen Sortierens von Marx & Hejj.

ANH 2.5
Gruppenmatrix der Distanz-Koeffizienten

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Bruder																
2	Cousin	759															
3	Cousine	858	451														
4	Enkel	777	770	866													
5	Enkelin	842	865	792	454												
6	Großmutter	860	882	830	702	570											
7	Großvater	783	822	877	564	695	444										
8	Mutter	747	942	895	836	803	679	834									
9	Neffe	787	538	674	802	882	925	865	970								
10	Nichte	900	685	542	875	783	869	926	927	463							
11	Onkel	819	594	711	804	875	856	772	913	592	719						
12	Schwester	432	857	764	841	781	774	853	673	884	797	892					
13	Sohn	632	892	953	721	869	864	764	710	900	969	912	743				
14	Tante	905	717	602	896	821	766	845	840	725	576	446	810	960			
15	Tochter	728	872	890	863	735	775	869	561	971	902	972	625	476	906		
16	Vater	614	896	947	787	894	823	693	477	915	981	831	744	571	924	718	

den Teilgruppen wieder in zwei Teile usw., bis jedes Wort an einem unteren Ende in der Hierarchie vereinzelt übrig blieb. Die Ähnlichkeit zweier Worte für einen Probanden wurde definiert als die Anzahl der Schritte, die nötig war, *um aus der isolierten untersten Position der beiden zu vergleichenden Wörter* in der Gruppierungshierarchie aufsteigend einen ersten gemeinsamen Knoten zu erreichen. Tabelle 2 enthält die über die Probanden summierten ‚Schritte bis zum ersten gemeinsamen Knoten'.

58 Diese lineare Transformation wurde lediglich zum leichteren Verständnis der aus der Tabelle abgelesenen Einzelwerte vorgenommen.

59 Eigenwerte: 4.31, 2.34, 1.54, 1.09, 1.05, 0.90, 0.79, 0.75…

Abbildung 2.08A zeigt die Varimin-Lösung. Der Generalfaktor F_1 hat nur minimale Ladungsvarianz, er ist im wesentlichen technisch-verfahrensbedingt und kann ignoriert werden.[60] Die Faktoren F_2 bis F_5 sind bipolar, sie lassen mit kontrastierenden Ladungsvorzeichen die erwarteten semantischen Merkmale erkennen: Linealität (F_2), Kernfamilie (F_3), Geschlecht (F_4) und Alter oder Generation (F_5). [61] Das minimale Wortpaar *Bruder* und *Schwester* z. B. zeigt einen Unterschied der Ladungsvorzeichen lediglich bei Faktor F_4, der als Geschlechtsfaktor (männlich vs. weiblich) zu deuten ist. Andere minimale Wortpaare lassen sich leicht finden, so z. B. *Vater* und *Sohn* oder *Mutter* und *Tochter*, die beide auf F_5, dem Faktor ‚Generation', mit ihrer Ladungsrichtung kontrastieren. In der Faktorenstruktur kommen zwar subtilere Unterschiede nicht zum Ausdruck, so etwa nicht die zwischen der jüngsten, mittleren und der ältesten Generation (etwa Sohn, Vater, Großvater). Auch werden Unterschiede, die durch die Ego-Perspektive bedingt sind, faktoriell ignoriert, z. B. die zwischen *(mein) Bruder* und *(meines Vaters) Sohn* – die gleiche Person ist gemeint. Doch sind die Hauptmerkmale der Verwandtschaft im Resultat aufzufinden.

---

[60] Man könnte F_1 als Ausdruck der Verwandtschaftskategorie deuten. Seine differenzierende Funktion würde faktoriell vermutlich erst in Erscheinung treten, wenn man in dieser Untersuchung zusätzlich Wörter einer anderen Kategorie verwendet hätte, die nichts mit Verwandtschaft zu tun hat.

[61] Die Interpretationen der Varimin-Faktoren F_2, F_4 und F_5 sind auch in einer vergleichbaren MDS-Untersuchung von Romney & d'Andrade (1964) anzutreffen (gender, generation, consanguinity). Die nunmehr vorliegende Untersuchung differenziert noch innerhalb F_2 (Linealität) zwischen ‚Kernfamilie' und ‚Nicht-Kernfamilie' (durch F3). Die Untersuchung von Romney & d'Andrade bietet ansonsten einen vorzüglichen Einstieg in die begrifflichen und methodischen Grundlagen der ‚componential analysis' von Begriffen und ein Verständnis für die Schlussfolgerungen, die man aus der Existenz der ‚discriminative stimuli' (der ‚Sememe') für den Begriffserwerb ziehen kann.

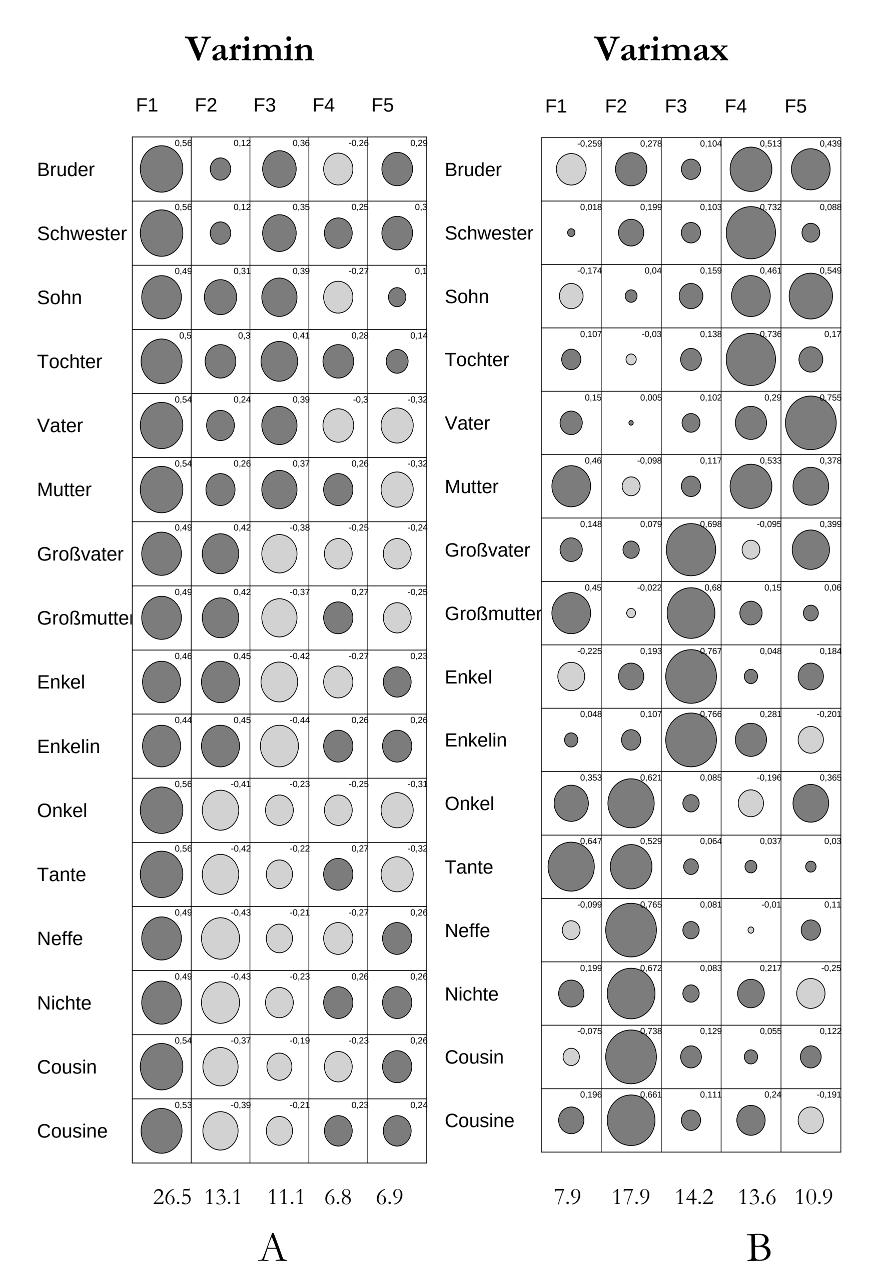


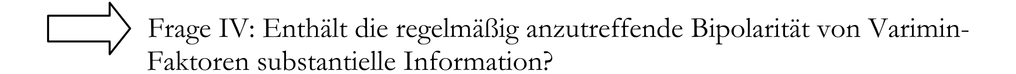
Abbildung 2.08

Varimin- (A) und Varimax-rotierte Faktoren (B)
der Analyse von Verwandtschaftsähnlichkeiten
(Daten von Marx & Hejj)

Abbildung 2.08B zeigt das Ergebnis einer *Varimax*-Rotation der Faktoren, die für Abbildung 2.08AVarimin-rotiert wurden. Von den fünf rotierten Faktoren lassen sich F_1 und F_5 gar nicht und F_3 nur mit einiger Willkür interpretieren. F_3 vereint Großeltern und Enkel, worin man einen Sinn hineinlegen kann, denn damit werden die generationsmäßig extremen Vertreter innerhalb derselben Linie gruppiert. Mit 'sehr jung oder sehr alt in der dominanten Abstammungslinie' würde man diese Einheit bezeichnen können.

Bei den Faktoren F_2 und F_4 hat das Varimax-Ergebnis eine gewisse Ähnlichkeit mit dem Varimin-Ergebnis. Mit Varimax-F_2 sind die Verwandtschaftswörter der *nicht-linealen* Abstammung geladen (‚entferntere Verwandte'). Im Unterschied zur Varimin-Lösung fehlt aber den Wörtern der *linealen* Abstammung (‚nähere Verwandte') als faktorielle Kennzeichnung der Vorzeichen-Gegensatz ([+],Merkmal vorhanden' vs. [–],Merkmal nicht vorhanden'), alle Varimax-Faktoren sind unipolar. F_4 gruppiert die Vertreter der Kernfamilie, doch die Nicht-Mitglieder bleiben bei F_4 wieder ohne Kennzeichen (kein negatives Vorzeichen). Außerdem haben *Enkelin* und *Nichte* und *Cousine* noch eine beträchtliche F_4-Ladung, womit man wenig anfangen kann.

Es ist also nicht so, dass sich Varimax-Faktoren einer Interpretation völlig entziehen. Doch wegen semantischer Überlappungen lässt ihre Interpretation meist zu wünschen übrig, was mithilfe der transparenten Merkmal-Semantik der Verwandtschaftswörter leicht zu erkennen ist. Vor allem bleiben im Varimax-Ergebnis auch bei den deutbaren Faktoren die Quellen der semantischen Ähnlichkeit und Verschiedenheit (Generation, Geschlecht usw.), die eine Faktorenanalyse aufdecken sollte, verborgen. Großeltern und Enkel werden aufgrund der Merkmale Linealität und Generation zusammen geclustert, doch werden diese Merkmale faktoriell nicht manifest.


Frage IV: Enthält die regelmäßig anzutreffende Bipolarität von Varimin-Faktoren substantielle Information?

Die Bipolarität der Ladungen in Varimin-Faktorlösungen verdient besondere Beachtung. Bipolarität tritt in Varimin-Lösungen regelmäßig auf, weit häufiger als in Varimax- und anderen SS-Lösungen. Das Vorkommen von negativen Ladungen kann bedeuten, dass die negativ geladenen Variablen gegenüber den positiv geladenen für die betreffende Quelle der Kovarianz (für den betreffenden Faktor) eine funktional abträgliche Wirkung haben. Bei den Response Set-Daten von Claus z. B. musste man bei den Variablen mit negativen Faktorgewichten eine hemmende Wirkung annehmen: Wenn jemand viele extreme Ja-Nein-Antworten abgibt, dann gibt er plausiblerweise weniger gemäßigte Antworten ab, und umgekehrt. Das lässt

sich seiner Motivation zuschreiben, die gemäßigten Urteile entweder zu meiden oder zu bevorzugen.[62] Für die bipolaren Akquieszenz- und Nuanciertheitsfaktoren gilt die motivational-funktionale Deutung des Vorzeichen-Kontrasts bei Faktor-Ladungen entsprechend.

Anders ist die Bipolarität bei den Verwandtschaftsdaten zu interpretieren. Beim Faktor Geschlecht z. B. kann Bipolarität nicht auch funktional gedeutet werden. Das Vorliegen des Merkmals *männlich* setzt keine Hemmung durch *weiblich*, oder umgekehrt voraus, sondern eine organismische Gegebenheit, die das Vorkommen des Merkmals *weiblich* ausschließt (in der Regel), wenn *männlich* vorliegt. [63] Das Merkmal Linealität setzt für sein Vorkommen oder Nichtvorkommen nur die richtige Antwort auf die Frage nach der entsprechenden Platzierung des Verwandten im Abstammungsbaum voraus, die ein *Ja* oder *Nein* ergibt. In solchen Fällen muss man sich darauf beschränken, mit den Vorzeichen das Vorkommen bzw. Nichtvorkommen eines Merkmals ausgedrückt zu finden. Dies wird auf anderen Gebieten, z. B. in linguistischen Diskursen, viel häufiger so gehandhabt als in der Psychologie – mit Recht so.

Der Fall, dass die negative Ladung eines Faktors das Nichtvorkommen eines Einflusses oder einer Funktion ausdrückt, kann darauf schließen lassen, dass der Faktor für die negativ geladenen Variablen irrelevant ist. Die Entstehung eines bipolaren Faktors ist in solchen Fällen anderen Variablen des untersuchten Begriffssamples zu verdanken, bei denen er positive Ladungen aufweist.[64] Man sollte sich bei CSM-orientierten Faktoranalysen darauf einstellen, negative Vorzeichen – wenn sie bei einem Faktor mit positiven zusammen vorkommen – zunächst immer als *‚Merkmal nicht vorhanden'* zu deuten und erst in einem zweiten Schritt, aufgrund eines von anderswoher bezogenen Kontextwissens, eine funktionale Deutung (Hemmung oder polar gerichtete Ausprägung usw.) zusätzlich einbringen, wenn eine solche nahe liegt.[65] [66]

---

[62] Es wäre denkbar, dass Probanden, die viele extreme Ablehnungen abgeben, viele Unentschieden-Urteile oder viele gemäßigte Zustimmungen abgeben, was motivational aber nicht nachvollziehbar wäre.

[63] Damit soll über die seelische Polarität von C. G. Jungs Animus vs. Anima, wonach die Tendenzen des Männlichen und Weiblichen innerhalb einer biologischen Geschlechtsausprägung auftreten, kein Urteil gefällt werden.

[64] Aus diesem Grund sind multivariate Kausalmodelle, wie sie mit den Methoden der konfirmatorischen Faktorenanalyse und den Strukturgleichungsverfahren Eingang fanden, als fragwürdig zu betrachten. Die Modellierung von Kovarianzquellen, die originäre Aufgabe der multivariaten Analyse, deckt nicht notwendigerweise kausale Quellen auf.

[65] Da die ‚Zuweisung' von positiven und negativen Vorzeichen durch das Rotationsprogramm vom zugeordneten Inhalt der Messungen unabhängig erfolgt,

Zum besseren Verständnis der Bipolarität von Faktoren ist es nützlich, ihr Entstehen anhand eines besonders transparenten Beispielfalls zu verfolgen:

Analyse 5: Intelligenzentwicklung in der Kindheit
(Daten von Humphreys & Davey).

In einer Längsschnittstudie testeten Humphreys & Davey (1988) Kinder vom dritten Lebensmonat bis zum 9. Lebensjahr mit insgesamt vier jeweils altersangepassten Intelligenztests. Sie wollten die Beständigkeit der intellektuellen Entwicklung in diesem Zeitraum ermitteln. Die von ihnen mitgeteilte Interkorrelationsmatrix für 14 Testwiederholungen im Längsschnitt wurde kopiert, PCA-faktorisiert und Varimin- und Varimax-rotiert. Abbildung 2.09 gibt die Ergebnisse wieder (die nicht-wiedergegebene Initiallösung ist mit der Variminlösung nahezu identisch). Nur zwei Faktoren sind substantiell (Eigenwerte: 5.705,1.473, 0.827, 0.788 …).

---

ist auch darüber zu entscheiden, ob die initial gegebenen Vorzeichen beibehalten oder der jeweiligen Deutung entsprechend ggf. umgekehrt werden sollten. Die Variablen, bei denen der Faktor funktional hemmenden Einfluss ausdrückt bzw. bei denen das faktoriell repräsentierte Merkmal als nicht vorhanden zu interpretieren ist, können zugunsten eines leichteren Verstehens ein negatives Vorzeichen erhalten, wenn dies nicht schon vorhanden ist. Es hat sich gezeigt (bislang eine Daumenregel), dass das Nichtvorliegen eines Merkmals oder der negative Einfluss durch eine Varianzquelle innerhalb eines faktoriellen Vektors tendenziell eher bei denjenigen Variablen vorkommt, (1) die niedrigere Ladungen haben und (2) deren Vorzeichen (Plus oder Minus) überwiegen .

[66] Ein aus der Frühzeit der Centroid-Faktorenanalyse stammender Hinweis, der den unrotierten bipolaren Faktoren einen Informationswert einräumt, der damals schon vernachlässigt wurde, findet sich in einer Notiz von Zimmermann (1953): *„It is well-known in dealing with intellectual variables that the first centroid loadings are usually all positive and the second centroid, as well as those that follow, divide positive and negative variables equally. What is apparently overlooked is the tendency for the second centroid to split the most obvious dichotomy, the third centroid to split the next most obvious dichotomy and so on. For example, if the battery contains both linguistic and quantitative tests, the second centroid will most likely separate these two major groups…“* (p. 389).



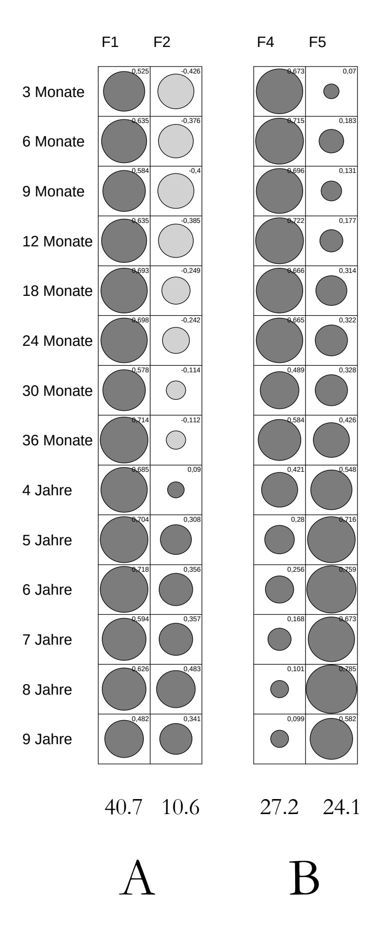


Abbildung 2.09

Varimin- (A) und Varimaxlösung (B) der Längsschnittdaten zur Intelligenzentwicklung, gewonnen von Humphreys und Davy.

Der erste Variminfaktor repräsentiert mit einer Kommunalität von 40.7% den stabil bleibenden Anteil der allgemeinen Intelligenz, die Faktorladungen bleiben innerhalb der Untersuchungszeit auf ungefähr gleichem Level. Der zweite Faktor ist bipolar, er repräsentiert mit 10.6% Kommunalität die zeitabhängige Varianz im Laufe der Intelligenzentwicklung. Diese ist plausiblerweise zurückzuführen auf günstige oder ungünstige Lebensumstände: auf Erziehungs- und Ausbildungseinflüsse, psychische oder körperliche Gesundheit oder Erkrankung usw., die unter den Testpersonen variieren.

Die Bipolarität von F_2 und die monotone Abfolge der Faktorgewichte werden wie folgt interpretiert: Der mit F_1 nicht ausgeschöpfte Kovarianzanteil verteilt sich ziemlich gleichmäßig über die Altersspanne, d. h. die Veränderungen von einem Messzeitpunkt zum nächsten sind geringfügiger als von einem Messzeitpunkt zum übernächsten. Bis zum *über-über*nächsten sind die Veränderungen noch größer, und vom ersten bis zum letzten Messzeitpunkt sind sie am größten. Dementsprechend korrrelieren die Testergebnisse der einander benachbarten Messzeitpunkte höher als die der einander entfernteren Messzeitpunkte. Die im mittleren Bereich des Längsschnitts liegenden Testwerte haben zu den Testanfangs- und -endwerten ungefähr gleiche Differenzen, die entsprechenden Korrelationen des Testergebnisses von der Mitte des Längsschnitts mit dem ersten und letzten Testergebnis sind ungefähr gleich hoch. Den Testzeitpunkt mit mittlerer Differenz gegenüber dem Anfangs- und Endwert macht das Varimin-Modell zum Nullpunkt des F_2-Ladungsvektors. Faktor F_2 repäsentiert somit im vorliegenden Fall das Ausmaß an Veränderung der Intelligenz gegenüber dem Mittelwert ihres stabilen Anteils.

Das Beispiel ist insofern lehrreich, als es zeigt, dass die Faktorgewichte von F_2 auf die von F_1 bezogen sind, wobei die Art des Zusammenhangs durch einschlägiges Kontextwissen spezifiziert werden muss. Negative F_2-Ladungen sind in diesem Fall weder Anzeichen von fehlendem oder gegensätzlichem Einfluss noch von logischem Ausschluss, sondern Anzeichen von Differenzen der Intelligenz-Veränderungswerte zum mittleren Veränderungswert. Da die F_2-Faktorgewichte positiv und negativ gepolt sind, entsteht eine Skala, die anzeigt, dass die Veränderung der Testergebnisse im Laufe der Längsschnitt-Untersuchung monoton zunimmt. Denn wenn z. B. ein genügend hoher konstanter Wert zu allen Faktorladungen hinzu addiert wird, kann die Bipolarität der Ladungen sowie der Nullpunkt verloren gehen, der Anstieg der Faktorladungen mit zunehmendem Alter bliebe erhalten, dieser würde nach wie vor die seit Beginn der Testperiode allmählich größer werdende Veränderung der Intelligenzleistungen repräsentieren. Die Varimaxlösung beseitigt den Generalfaktor und ignoriert somit die Tatsache, dass die Intelligenz im getesteten Längsschnitt relativ konstant bleibt. Stattdessen wartet sie mit zwei Faktoren auf, die auf den ersten Blick wie zwei unabhängige Intelligenzen erscheinen, die eine, die sich in der frühesten Kindheit auswirkt, während eine andere Intelligenz die früher auftretende allmählich ablöst – eine absurde Vorstellung, zu der auch SSM-Faktoristen kaum neigen werden. Sie würden ihr Verfahren im Falle von Längsschnitt-Wiederholungsdaten vermutlich nur für ungeeignet ansehen und auf seine Anwendung verzichten.

Ein andersartiges Beispiel, bei dem die negative Ausprägung eines Faktors nicht lediglich Abwesenheit eines Merkmals, sondern ein skalometrisch eigenständiges Merkmal erkennen lässt, bietet die folgende Untersuchung.

Analyse 6: Körpervolumen und Körperform von Rindern
(Daten von Rasch).

Von 107 weiblichen Rindern (Färsen) erhob Rasch (1962) 12 Maße ihrer Körperausdehnung. Es handelte sich um Höhen-, Breiten- und Längenmaße. Unsere übliche faktorielle Verarbeitung der von E. Weber mitgeteilten Interkorrelationsmatrix führte zu der in Abbildung 2.10 wiedergegebenen zweifaktoriellen Lösung (Eigenwerte: 7.69, 1.20, 0.74…). Die nicht wiedergegebene Initialstruktur ist mit der Variminstruktur nahezu identisch.

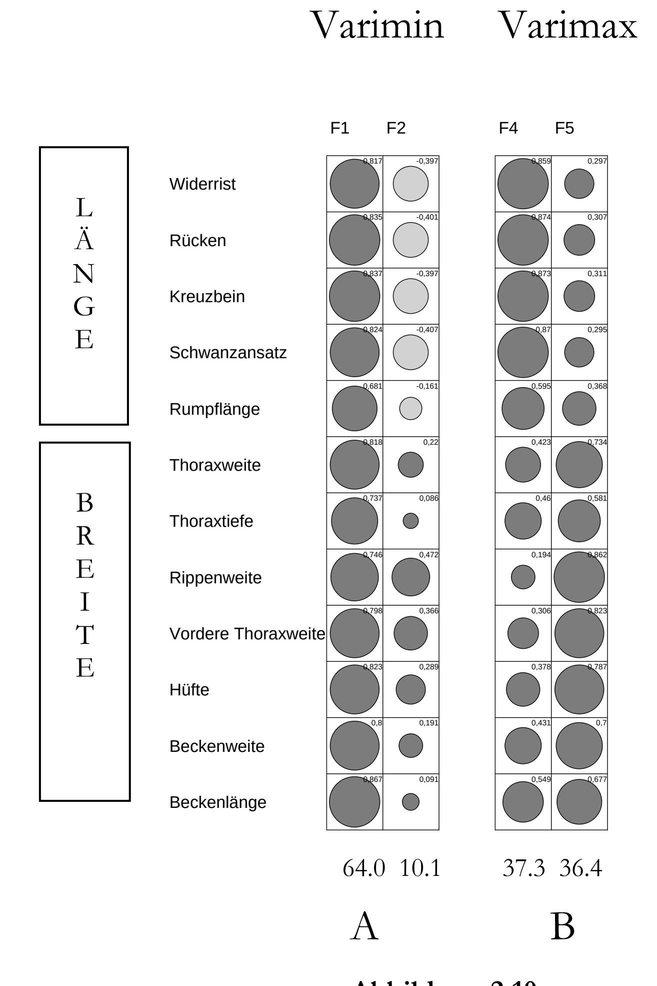


Abbildung 2.10
Varimin- (A) und Varimaxfaktor (B) von
12 Körpermaßen weiblicher Rinder (Färsen)

Der erste Faktor (Kommunalität 64.0%) repräsentiert das Körpervolumen (die Körperausdehnung) ohne Spezifizierung einer Raumdimension. Er besagt, dass ein Rind, das über- oder unterdurchschnittlich hoch gewachsen ist, meist auch über- bzw. unterdurchschnittlich in die Länge und Breite gewachsen ist. F_1 ist dem Generalfaktor der Intelligenz analog zu deuten.

Der zweite Faktor (Kommunalität 10.1%) repräsentiert mit seiner Bipolarität die Varianz der Körperform. Es gibt unter sehr voluminösen wie unter wenig voluminösen Rindern relativ schmal gewachsene (schlanke) und relativ breit gewachsene (dicke). Das Auftreten einer Polarität zeigt an, dass mit mäßigem Spielraum das Wachstum in die Länge ‚auf Kosten der Breite' und das in die Breite ‚auf Kosten der Länge' vonstattengeht.

In der Varimaxlösung geht der Generalfaktor des Körpervolumens verloren. Die Variablen der Schlankheit und Dicke sind zwar auch erkennbar, allerdings als nicht-polare orthogonale Dimensionen. Die empirische Gegenläufigkeit der Längen-und Breitenausdehnung, die sich durch Varimin aufgrund der Dominanz des Volumenfaktors zu einer bipolaren Einflussgröße der ‚Körperform' herausbildet, kommt im SSM-Ergebnis nicht zum Tragen. [67]

Die Deutung negativer Ladungen eines Faktors als Ausdruck eines bipolar aufzufassenden Charakteristikums kann auch bei psychologischen Daten naheliegen, demonstrierbar mit folgendem Beitrag.

[67] Die Anwendung eines SSM auf psychologische Daten hat vielfach zur Auflösung bipolarer Konstrukte geführt: Urteile über negativ getönte Gefühle erweisen sich nach Anwendung des SSM als statistisch unabhängig von Urteilen über positiv getönte Gefühle, man glaubt dann entgegen aller Lebenserfahrung, sie seien funktional voneinander unabhängig (Diener & Emmons, 1983, Watson & Clark, 1988, späte Korrektur nach Modellvergleichen durch Crawford & Henry, 2004). Ähnlich zusammenhangslos erscheinen in SSM-Faktorenanalysen Optimismus und Pessimismus, die aufgrund universeller Erfahrung bipolar (gegensätzlich) sich auswirkenden Einstellungen zur Zukunft (Marshall et al., 1992). Auch geht die Polarität der Geschlechter-Typologie durch Faktorenanalysen verloren: Femininität soll allein aufgrund faktorieller Orthogonalität nicht auf Maskulinität polar bezogen sein (Bem 1981).

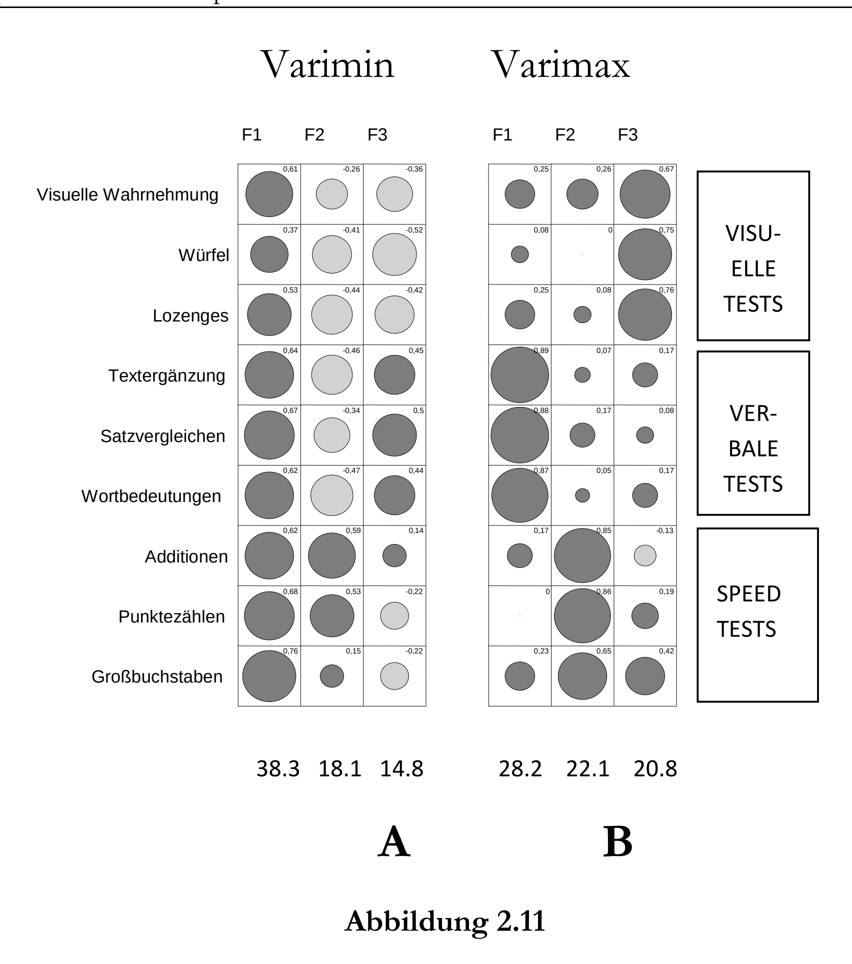


Abbildung 2.11

Varimin- (A) und Varimaxlösung (B) der Holzinger & Swineford-Daten

Analyse 7: Intelligenz- und Leistungstests (Daten von Holzinger & Swineford).

Die Holzinger & Swineford-Daten werden gelegentlich in Lehrtexten zur Demonstration von Modellrechnungen verwendet, so auch von Jöreskog & Sörbom (2003), von deren Website die Interkorrelationstabelle entnommen wurde (zu ergoogeln mit ‚LISREL 8.52 Jöreskog'). Es handelt sich um je drei Tests für visuelle, für verbale und für Tempo-Leistungen (speed tests). Die Ergebnisse der vorliegenden Standard-Auswertung für Varimin und Varimax zeigt Abbildung 2.11.

Varimin-F_1 (Abbildung 2.11A) lässt sich als Generalfaktor ‚g' der allgemeinen Intelligenz ansehen, der für Intelligenztestbatterien generell zu erwarten ist. Der

Faktor F_2 kontrastiert durch gegensätzliche Vorzeichen die drei Speed-Test-Leistungen mit den übrigen Testleistungen, bei denen es offensichtlich weniger auf Tempo als auf Konzentration oder ‚power' ankommt. Hier liegt ein Fall vor für Bipolarität im Intelligenzbereich, bei dem – mit psychologischen Kontextkenntnissen – negative Ladungen mit funktionaler Bedeutung in Zusammenhang gebracht werden dürfen. Eine erhöhte Fähigkeit und Neigung zu Tempoleistungen wirkt sich vor allem bei geeigneten Speed-Tests förderlich aus. Bei den Tests, die Konzentration und Vertiefung in Probleme erfordern, werden sich Speed-Neigungen und -Fähigkeiten vermutlich eher nachteilig auswirken. Eine entsprechende Gegenwirkung lässt sich auch für die Fähigkeit und Neigung zur Konzentration denken, die sich bei Tests, die Tempofähigkeit und -neigung erfordern, vermutlich nachteilig auswirken. Diese Gegenläufigkeit kann sich mit der Bipolarität von F_2 manifestieren.

Die Bipolarität von Varimin-F_3 in dieser Untersuchung hebt eine andere Kovarianzquelle ans Licht: Die Besonderheit verbaler Testleistungen. Bemerkenswert ist, dass sich diese Bipolarität nicht auf verbale vs. visuelle Tests beschränkt, die beide zur Power-Kategorie gehören. Wegen der unterschiedlichen Aufgabeninhalte (Wörter vs. Bildmaterial) bilden verbale und visuelle Tests gute minimale Kontraste. Doch auch die rechnerischen Leistungen aus der Speed-Testserie kontrastieren, wenn auch etwas schwächer, mit den verbalen Testleistungen. Offenbar liegen den verbalen Leistungen übergreifende Besonderheiten zugrunde. Auch hier lässt sich eine antagonistische Beziehung denken: Die verbalen Fähigkeiten und Neigungen, wenn sie im individuellen Intelligenzprofil dominieren, könnten die rechnerischen als auch die visuellen Aktivitäten etwas benachteiligen und/oder umgekehrt. Doch kann die funktionale Relevanz bei negativen Varianzquell-Ladungen hier wie oft in anderen ähnlichen Fällen nur Hypothese sein. [68]

Die Varimax-Lösung (Abbildung 2.11 B) bringt ‚g' zum Verschwinden und separiert die drei Testsorten voneinander, so dass sich mit dieser Lösung kein Antagonismus zwischen verschiedenen Fähigkeiten und Neigungen manifestieren kann. Die LISREL-Lösung von Jöreskog & Sörbom kommt mit größerem Aufwand zum gleichen Ergebnis wie die Varimax-Lösung von 11 B. Die Autoren hatten das SSM-Modell mit den drei latenten Varianzquellen, deren Unabhängigkeit vorausgesetzt wurde, konfirmatorisch überprüfen wollen und waren dabei auch erfolgreich – allerdings auf Kosten der naheliegenderen und theoretisch plausibleren

[68] Man könnte die Tests mit zwei alternativen Zusatzinstruktionen durchführen lassen, eine Speed-Instruktion würde die Probanden generell zu mehr Tempo ermuntern, eine Power-Instruktion zu mehr bedachtsamer Konzentration. Vermutlich wird man je nach Testart differenzielle Steigerungen bzw. Verminderungen ihrer Leistung beobachten.

Lösung, zu der Varimin geführt hat. Dass in bipolaren Lösungen nicht immer Variablen nur entweder mit positiven oder negativen Ladungen vorkommen, sondern dass sie auch Nahe-Null-Ladungen haben können, und dass dies informativ sein kann, zeigt das folgende Beispiel.

Analyse 8: Psychophysiologische Aktivationsindikatoren
(Daten von Köhler & Troester).

Den Autoren Köhler & Troester (1991) ging es um die Validierung des Maßes der Handflächen-Schweißproduktion (PSI, Palmar Sweat Index) als Indikator für psychophysiologische Aktivation. Sie testeten 50 Personen unter drei Ruhebedingungen und einer Anstrengungsbedingung (von der Zahl 2007 hatten die Probanden sukzessive die Zahl 7 zu subtrahieren) und erhoben psychophysiologische Werte pro Person 16 mal im Laufe der vier Testphasen, nämlich PSI-Werte am Mittel- und Zeigefinger (PSI-M, PSI-F), spontane Fluktuationen der Schweißproduktion (SF), den Hautwiderstand (SCL, skin conductance level) und die Herzschlagfrequenz (HR). Die fünf erhobenen Werte wurden pro Person über die 16 Messwiederholungen intraindividuell korreliert, die Korrelationen wurden über die 50 Probanden gemittelt. Die Ergebnisse unserer Standardauswertung der faktorisierten Interkorrelationen liefert Abbildung 2.12. Zwei Faktoren wurden nach dem Eigenwertverlauf 3.50, 0.76, 0.33, 0.20 …und nach dem Ladungsmuster für interpretierbar gehalten, auf die Wiedergabe der Initiallösung, die der Variminlösung sehr ähnlich ist, wird verzichtet.

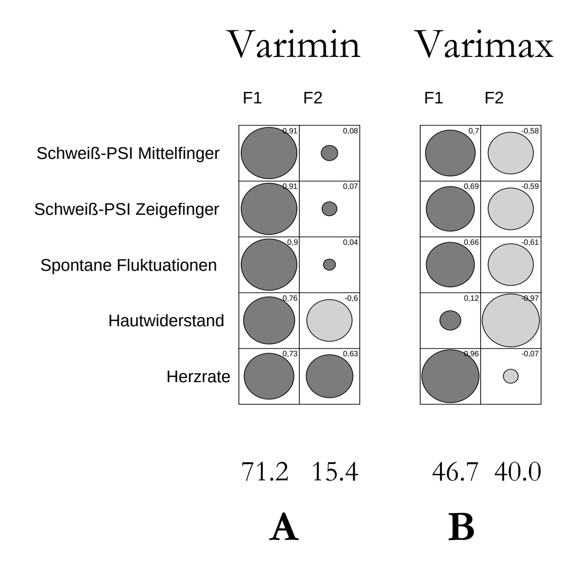


Abbildung 2.12
Varimin- (A) und Varimaxlösung (B)
der Köhler- und Tröster-Daten

Ergebnisse: Varimin-F_1 ist ein Generalfaktor, der besagt, dass den individuellen Differenzen der fünf Maße offensichtlich der gleiche Indikatorwert – Aktivation – zukommt. Varimin-F_2 repräsentiert Zusatzvarianz, die sich allerdings auf den Hautwiderstand und die Herzrate beschränkt, die Schweißvariablen haben mit dieser Varianzquelle offenbar nichts zu tun, dies zeigen ihre Um-Null-Ladungen an. Die F_2-Ladungen mit gegensätzlichen Vorzeichen bei SCL und HR besagen, dass die beiden Variablen offensichtlich ein leicht antagonistisches funktionales Verhältnis zueinander haben: Es gibt Probanden, bei denen der Hautwiderstand auf Aktivation stärker reagiert als die Herzfrequenz und andere, bei denen umgekehrt die Herzfrequenz stärker reagiert als der Hautwiderstand. Obgleich beide Variablen mit F_1 im wesentlichen die gleiche Aktivationsfunktion indizieren, bleibt ein geringer Rest an Varianz übrig, der durch eine leichte Präferenz entweder des Hautwiderstands oder der Herzrate aufgeklärt wird. Hier scheint ein Anteil ‚gegabelter Wirkung' vorzuliegen: Je mehr Wirkung in Richtung X vorliegt, umso weniger Wirkung ergibt sich für Richtung Y, und umgekehrt. Die Schweißproduktion ist von diesen koordinierten Präferenzwirkungen nicht betroffen.

Die Varimaxlösung stellt den Auswerter vor zwei unlösbare Rätsel. Die Aufspaltung des Aktivationseffekts auf zwei unabhängige Äste ist das eine Rätsel – aus physiologischen Gründen sind zwei unabhängige Aktivationsquellen kaum vorstellbar. Das zweite Rätsel ist die Zugehörigkeit von SRL zum einen Aktivationsast und von HR zum anderen Aktivationsast. Varimax vermischt hier die Varianzquellen bis zur Unkenntlichkeit.

Die bisherigen Beispiele zeigten: Die faktorenanalytische Auswertung von Variablen, die der hypothetischen Komplexität ihrer latenten Varianzquellen gerecht werden will, führt zu Ergebnissen, die leichter interpretiert werden können als Ergebnisse, denen man eine Einfachstruktur aufzuzwingen sucht. Die Ebene der latenten Bedingungen wird im Folgenden differenzierter in den Blick genommen.

⇨ Frage V: Sind mithilfe der CSM-orientierten Faktorenanalyse auch methodenbedingte Einflüsse erfassbar?

Campbell & Fiske legten 1959 einen neuen methodologischen Ansatz vor, der in den darauf folgenden Jahrzehnten die Entwicklung zahlreicher methodischer Verfahren unter dem Label MTMM (multi trait multi method) nach sich zog. Die Autoren nahmen systematisch in Angriff, worauf man schon vorher, aber weniger nachdrücklich, aufmerksam gemacht hatte: Persönlichkeitsforscher haben nicht nur mit einer Vielheit von Varianz-erzeugenden latenten Traits zu tun. Mit zusätzlicher Varianz haben sie zu rechnen, wenn sie zu deren Erfassung unterschiedliche Testverfahren verwenden. Bald darauf bezog man auch die Varianzanteile mit ein,

die durch unterschiedliche Informanten (Selbst- vs. Fremdbeurteilung) entstehen. Andere mögliche Varianzeinflüsse, die an den Kovarianzen der gemessenen Variablen einen Anteil haben können, kamen hinzu (Situationseinflüsse bei Testwiederholungen usw.). Da sich das hier favorisierte komplexitätsoffene Verfahren beim Aufspüren multipler Varianzquellen bewährt hat, liegt es nahe, eine entsprechende Leistung auch für MTMM-Daten zu erwarten, ohne dass ein Wechsel des Analyseverfahrens notwendig wäre.

Analyse 9: Wissensprüfung mit Variation der Prüfmethode
(Daten von Campbell & Fiske).

Zur Differenzierung methodischer Varianzquellen entwickelten Campbell & Fiske (1959) ein Verfahren, das auf eine systematische Inspektion der Interkorrelationstabellen hinausläuft. Sie demonstrierten ihr Vorgehen mit Korrelationsdaten, die sie in der Fachliteratur fanden. Ein Datensatz aus Campbell und Fiske, der schon von Cronbach und Vernon für andere Zwecke reanalysiert worden war – ihre Quelle geben sie nicht an – stammt von einem Wissenstest, bei dem im Physik-Unterricht vermittelte Inhalte zur Mechanik und Elektrizität durch verbale Fragen sowie durch Fragen mithilfe bildlicher Darstellungen geprüft wurden.

Die Ergebnisse unserer Auswertung für Varimin und Varimax zeigt Abbildung 2.13. Nach dem Eigenwertverlauf (2.67, 0.62, 0.45, 0.27) wäre nach üblichen Kriterien nur F_1 , allenfalls noch F_2 heranzuziehen. Doch Campbell & Fiske erkannten mithilfe ihrer Korrelationsinspektion recht deutlich die Auswirkung auch der beiden Varianzquellen, die sich in unserer Faktorenanalyse als F_2 und F_3 auswirkten. [69]

[69] Die Eigenwertkriterien der Mindesthöhe (>1) bzw. der Eigenwertknick (zu ermitteln durch den Scree Plot) sind schon von anderer Seite als unzuverlässig erkannt worden (Fabrigar, 1999, p.287). Der u.a. auch von Fabrigar empfohlene ‚Paralleltest' ist mit den vorliegenden Korrelationsdaten nicht durchführbar. Nach eigenen Erfahrungen mit faktoriellen Reanalysen von Datensätzen haben sich hier als hilfreich erwiesen die Prozentanteile, die die Initialfaktoren an der Summe der Kommunalitäten haben. Ein Faktor, dessen Anteil an der Summe h^2 weniger als 10% beträgt, ist in der Regel nicht mehr interpretierbar. Im vorliegenden Datensatz von Campbel & Fiske betragen die Kommunalitätsprozente der vier Initialfaktoren 66.7, 15.4, 11.3, 6.7. Wenn ein Mindestanteil von 10% Kommunalität als Kriterium angesetzt wird, sind F_2 und F_3 noch zu berücksichtigen.

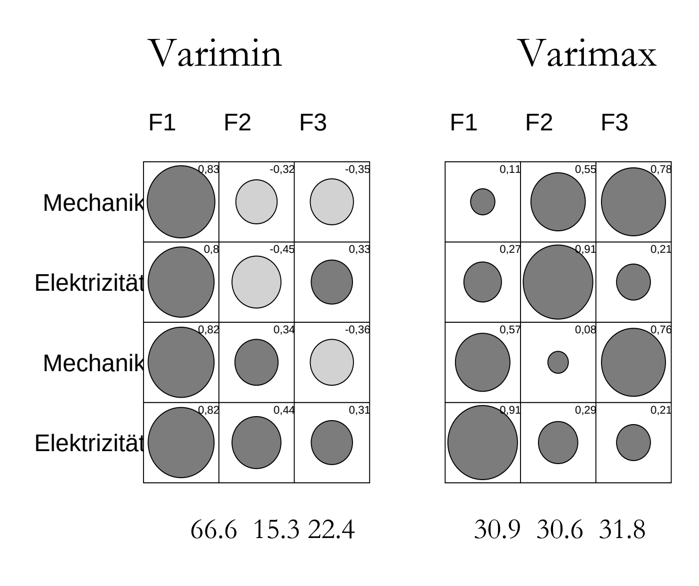


Abbildung 2.13

Varimin- (A) und Varimaxlösung (B)
der Campbell & Fiske-Daten

Ergebnis: Varimin-F_1 repräsentiert mit hohem Kommunalitätsanteil den Generalfaktor, zu dem wohl nicht nur die unterschiedliche Intelligenz der Probanden, sondern auch ihr unterschiedlicher Fleiß für das Unterrichtsfach Physik beigetragen haben. Hinzu kommt mit weitaus geringerem Gewicht F_2 als Methodenfaktor (Sprache vs. Bilder) und F_3 als Faktor der Lehrinhalte (Elektrizität vs. Mechanik). Unter den Probanden gab es offenbar solche, die mehr mit verbal ausformulierten Fragen als mit Abbildungen zurechtkamen und solche, denen Probleme mit Abbildungen leichter fielen (F_2). Auch hatten einige wohl mehr für die Mechanik als für die Elektrizität gelernt und andere umgekehrt (F_3). Wieder erweist sich hier Bipolarität als ein Indikator für konkurrierende Gegebenheiten, nicht etwa nur für das Vorliegen oder Nicht-Vorliegen einer einzigen Gegebenheit.

Die Gegenläufigkeit der Wirkungen bei verbalem vs. bildlichem Prüfmaterial geht in der Varimaxlösung verloren. Ferner fehlt in der Varimaxlösung eine faktorielle Repräsentanz der Elektrizitätsinhalte, nur die Mechanikinhalte werden durch Varimax faktoriell repräsentiert. In einem letzten Beitrag soll die Frage geklärt werden, ob sich durch Varimin auch die Varianzquelle ausdrücken wird, die durch den Wechsel befragter Personenstichproben entsteht.

Analyse 10: Selbst- und Fremdbeurteilung von Kindern
(Daten von Matson & Nieminen)

In einer Fragebogen-Untersuchung zu Verhaltensstörungen, Depression und Angst bei Kindern von Matson & Nieminen (1987) verwendeten die Autoren sechs Skalen, die den Kindern selbst sowie ihren Lehrern/Lehrerinnen vorgelegt wurden, welche die Kinder zu beurteilen hatten. Das Ergebnis unserer Auswertung zeigt Abbildung 2.14. Die Eigenwerte betragen 3.98, 2.14, 1.41, 0.99, 0.86…, die Prozentanteile an aufgeklärter Varianz bei den Initialfaktoren sind 33.1, 17.8, 11.8, 8.2…, wonach drei Faktoren als substantiell zu bezeichnen sind.

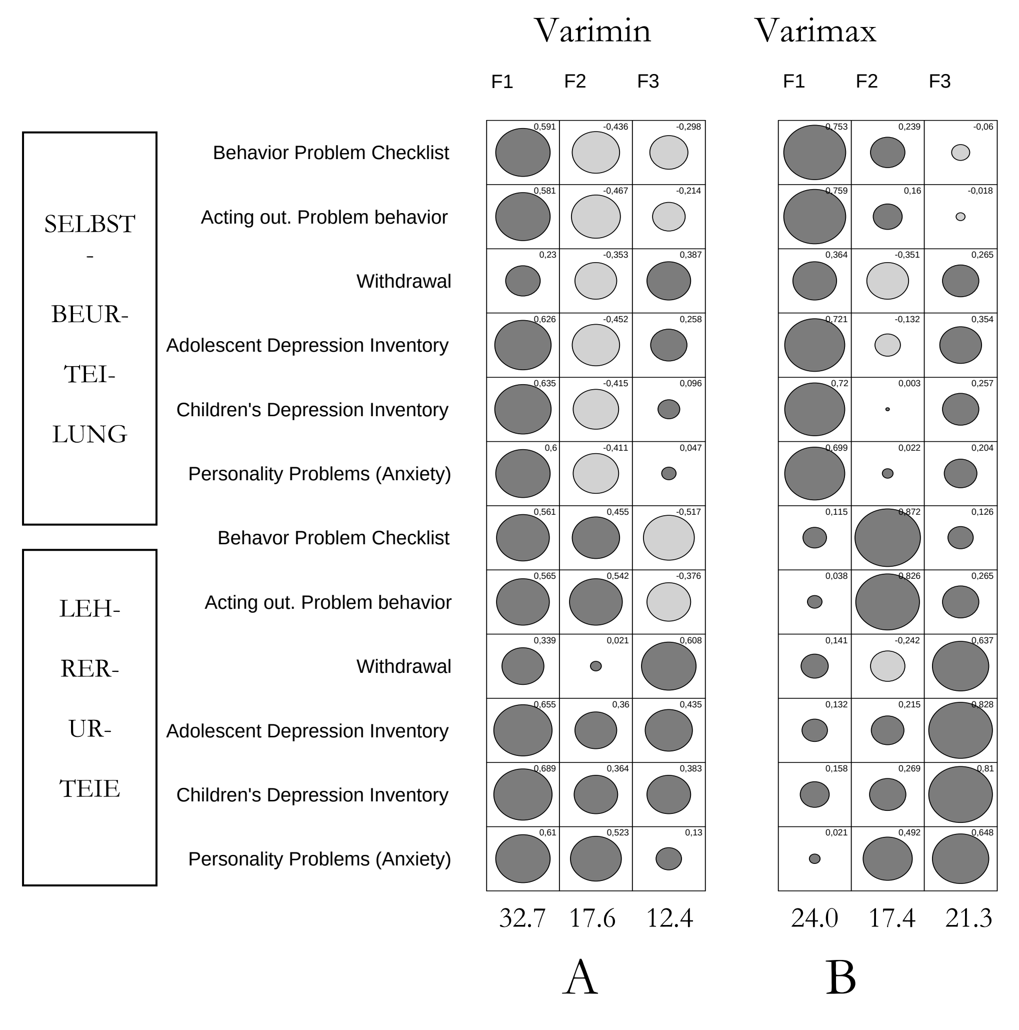


Abbildung 2.14

Varimin-(A) und Varimaxlösung (B)
der Matson & Nieminen-Daten

Zum Ergebnis der *Varimin*lösung, Abbildung 2.14A: Die im Generalfaktor F_1 sich ausdrückende Gemeinsamkeit der Kovarianzquelle bei Kindern und Lehrern lässt mit etwas Zurückhaltung darauf schließen, dass in den Urteilen tatsächlich vorliegende dysfunktionale Symptome bei den Kindern mit ihrer interindividuellen Varianz zum Ausdruck gebracht werden. Zurückhaltung ist geboten, weil sich bei den meisten Fragebogenuntersuchungen, so auch wohl in dieser, im ersten Faktor zusätzlich die Akquieszenz als Beantwortungsstil geltend machen kann. Dass in dieser Untersuchung die Variablen mit F_1 bei Kindern und Lehrern unterschiedslos geladen sind, ist kein stichhaltiger Einwand, da Akquieszenz (Ja-Sage-Tendenz) als Urteilsgewohnheit bei Kindern und Lehrern gleichermaßen variieren kann. Zusätzlich kann zu F_1 bei Kindern *und* Lehrern in gleicher oder ähnlicher Weise ein Einfluss der *social desirability* (SD) beigetragen haben, der nicht sicher erkennbar ist, da alle Skalen negativ bewertetes Erleben und Verhalten zum Inhalt haben.

Mit gegensätzlichen Vorzeichen beim bipolaren F_2-Faktor macht sich offenbar der Informanten-Einfluss geltend. Dieser ist als solcher zwar nicht zweifelhaft, seine Herkunft aber bleibt mehrdeutig. Es mag sein, dass die Kinder sich selbst tendenziell als weniger erlebens- und verhaltensgestört beurteilen als die Lehrer, oder auch als gestörter, oder dass sie zu mehr oder zu weniger Akquieszenz oder SD neigen als die Lehrer. Auch können sich diese Einflüsse ähnlich wie die Teil-Einflüsse bei F_1 addieren.

Mit dem Vorzeichen-Gegensatz des bipolaren Faktors F_3 werden die dysfunktionalen Symptome inhaltlich differenziert, bei Kindern und Lehrern in gleicher Weise. Auf der einen Seite gruppieren sich die Störungen mit enthemmender Auswirkung (acting out, conduct disorder), auf der anderen Seite die Störungen mit gehemmter Wirkung (withdrawal, depression, anxiety). Auch hier ist eine antagonistische Beziehung anzunehmen, dysfunktionale Dispositionen können sich je nach Veranlagung entweder als Hemmungen ausdrücken oder aber zu einem enthemmenden ‚acting out' führen (Aggressionen z. B.), wobei Angst ‚abgeführt' wird, was psychoanalytisch so gedeutet werden könnte.

Die *Varimax*lösung (Abbildung 2.14 B) führt zu drei Clustern der Schüler- und Lehrerskalen. In einem Cluster sind die sechs von den Schülern beurteilten Skalen vereint, in einem zweiten die drei von den Lehrern beurteilten Skalen der Enthemmungsstörung, in einem dritten Cluster die zwei von den Lehrern beurteilten Skalen der Hemmungsstörung, wobei eine Reihe von Ausreißern und mehrfaktorielle Ladungen das Bild beeinträchtigen. Mit diesen Clustern lässt sich konzeptuell kaum etwas anfangen.

Unsere Reanalyse der Daten von Matson & Nieminen hat nochmals deutlich gemacht, dass der Ausdruck ‚latent' im Sinne von ‚unsichtbar', ‚verborgen' zur Kennzeichnung der Faktoren, die den Variablen zugrunde liegen, nicht immer

zutrifft. Die Verwendung wechselnder Beurteiler liegt als Bedingungsfaktor offen zutage. Ihr Einfluss, wenn ein solcher vorliegt, sollte sich durch eine Analyse, welche Varianzquellen aufzudecken hat, ebenso zeigen wie der Einfluss zunächst unbekannter Varianzquellen, oder auch nicht zeigen, wenn kein solcher Einfluss vorliegt. Im vorliegenden Fall konnte die Interpretation von Varimin-F_2 (Informanten) auf einen transparenten Einfluss zurückgreifen, die Interpretation von F_3 (Enthemmungs- vs. Hemmungsstörung) musste sich mit einer weniger transparenten psychologischen Unterscheidung und Abstraktion begnügen.

Diskussion zu Kapitel 2

Fünf Fragen zur Methodologie wurden behandelt, die sich mit der Verwendung des neuen Rotationsverfahrens Varimin ergeben. Zehn empirische Anwendungsbeispiele sollten helfen, sie zu klären. Enthielt die Auswahl der zehn Beiträge zu wenig Daten psychologischer Herkunft? Die Güte einer neuen Methode sollte zunächst mit Daten geprüft werden, bei denen zweifelsfrei bestimmte Ergebnisse zu erzielen sind, sofern die Methode funktioniert. Wenn ein in jedem Fall zu erwartendes Ergebnis nicht erzielt wird, soll man schlussfolgern dürfen, dass mit der Methode etwas nicht stimmt. Von dieser Prüfstrategie machen Methodiker unter den Psychologen m. E. generell zu wenig Gebrauch. Psychologische, sprachliche Daten insbesondere, eignen sich zur Methodenprüfung kaum. Die Sprache hat zwar zur Differenzierung von Erlebens- und Verhaltensweisen eine Fülle an Wörtern und Wendungen hervorgebracht, ihr Informationswert ist beträchtlich. Doch ist die Semantik der Psychologie weit weniger transparent als z. B. die Semantik der Verwandtschaft.

Natürlich soll die Faktorenanalyse, wenn sie ihre methodischen Prüfungen bestanden hat, bei der Lösung inhaltlicher Probleme unserer Disziplin eingesetzt werden. Aus den inzwischen vorliegenden Ergebnissen der Methodenforschung lässt sich schon jetzt eine bedeutsame Schlussfolgerung voraussagen: Auch dem manifesten verbalen Urteilsmaterial, das Psychologen zur Analyse einbringen, liegen auf einer latenter Ebene Komponenten- oder Merkmalstrukturen zugrunde. Diese sind indessen viel schwerer zu identifizieren und voneinander zu separieren als die Merkmale, die in unserer anschaulich gegliederten Lebenswelt zu finden sind. [70] Deshalb wird man sich bei der Interpretation von Varimin-Faktoren einer

[70] *"Characterizing concepts in terms of features works very well for certain types of words …[those] that are very well structured by physical, social, or biological dimensions, which then serve as the basis of semantic features. Other types of words… which do not come from well-structured domains, are more difficult to characterize as a set of features. …No metho-*

psychologischen Domäne nicht, wie gewöhnlich bei der Interpretation von SS-Faktoren, mit bloßen Faktor-Benennungen begnügen dürfen, die sich an Bezeichnungen der manifesten Variablen (Fragebogen-Items z. B.) anlehnen. Man wird sich auf die Ebene der latenten Merkmale begeben und sich trennen müssen von Begriffen umgangssprachlicher Herkunft, z.B. von Trait-Konzepten wie *agreeable, open, sociable, conscientious,* usw., weil nicht diese selbst, sondern deren Varianzquellen gefragt sind. Man wird sich ähnlich verhalten müssen wie bei der Analyse der Verwandtschaftsdomäne, bei der man zur Entschlüsselung der Faktoren auf eine latente Ebene hinabsteigen und sich von den manifesten Begriffen *Mutter, Schwester, Tante* usw. trennen musste, weil nicht diese Begriffe selbst oder Cluster von ihnen, sondern deren *distinctive features* gefragt waren.

Die empirischen Beispielfälle, bei denen relativ transparente Domänen von Variablen einer Analyse unterzogen wurden, haben gezeigt, dass man Varimin-Faktoren durchaus deuten kann. Die Deutung der Variminlösungen war sogar weit befriedigender als die der Varimaxlösungen. Mit Anwendung des minimalen Paarvergleichs, der bei schwierigen Variminfaktor-Deutungen vorgenommen werden kann, beschränkt man sich auf Unterschiede zwischen gepaarten Variablen hinsichtlich nur eines Merkmals, was durchweg leichter fällt als das Auffinden von Gemeinsamkeiten bei unanalysierten Merkmalkomplexen, wie sie durch Varimax geclustert werden.

Drei Einschränkungen sind indes nachzutragen. Erstens, die Anwendung des Minimalpaar-Vergleichs setzt voraus, dass im Datensatz geeignete Variablenpaare vorhanden sind. Oft gewinnt man sichere Deutungen eines Faktors erst, wenn mehrere Minimalpaare vorliegen, die bei gleichen Merkmalkontrasten auch paarweise gleiche Merkmalprofile haben, die aber von Paar zu Paar variieren dürfen. Da man dieses Desiderat bei der Variablenselektion im Voraus oft nicht berücksichtigen kann, wird ein Faktor mitunter nicht sicher gedeutet werden können, weil minimale Paare fehlen. Die noch fehlenden Paare könnten sich erst mit einer repräsentativeren Variablen-Stichprobe einstellen.

Das Vorkommen minimaler Paare ist umso eingeschränkter, je weniger Variablen verwendet wurden und je mehr substantielle Faktoren extrahiert wurden, so dass der Bedarf an Kombination gegensätzlicher Faktorgewichte zur Bildung von Minimalpaaren mitunter nicht gedeckt ist. Deshalb empfiehlt es sich, bei der Zusammenstellung von Variablenstichproben für eine faktorielle Untersuchung einer Domäne generell die Zahl der vermuteten Varianzquellen eher etwas zu beschrän-

dology directly reveals the meaning components of a word. Despite the difficulty in confirming what is or is not a feature of a word, it is generally accepted that … words are comprehended by accessing various features or meaning components." (1984, Just & Carpenter, p. 63 f).

ken, aber innerhalb des gewählten Vermutungsbereichs wechselnde Items großzügig zuzulassen.
Die zweite Erschwernis resultiert aus dem Umstand, dass die Varianzquellen, die mit Varimin gefunden werden, immer einen höheren Grad an Abstraktheit besitzen als die Begriffe, die man zur Beschreibung von SS-Faktoren verwendet. Mit den Verwandtschaftswörtern lässt sich das andeuten: Das Merkmal *Generation*, mit Varimin gewonnen, ist abstrakter als ein Begriff wie *die Jüngsten und Ältesten in der Abstammungslinie*, der bei der Varimax-Analyse notwendig war, um Großeltern und Enkel unter einen faktoriellen Hut zu bringen. Man wird in der psychologischen Domäne mit der Identifizierung abstrakter Merkmale größere konzeptuelle Anstrengungen aufbringen müssen als mit konkret typisierenden Begriffen, wie sie in der SS-Praxis verwendet werden. Auch unter *Möbeln* z. B. kann man sich nicht so schnell etwas Konkretes vorstellen als unter *Kleiderschrank* oder *Polstersessel.* [71]

Schließlich ist die Minimalpaarbildung in der Regel bei Generalfaktoren nicht anwendbar. Die Generalfaktor-Ladungen der Variablen einer Domäne sind meistens durchweg positiv, und sie weisen meist wenig Varianz auf. Andererseits repräsentiert ein Generalfaktor in der Regel die für den Datensatz jeweils dominante Varianzquelle, an der ein Forscher in erster Linie interessiert ist und die naheliegender ist als hinzu kommende weniger gewichtige Varianzquellen. Man wird einen Generalfaktor bei der Analyse von z. B. psychophysiologischen Variablen selbstverständlich auf z. B. Aktivationsvarianz zurückführen und nicht etwa auf intellektuelle Varianz.

Vor allem bei Generalfaktoren ist immer auch daran zu denken, dass sich mit ihnen mehr als nur *eine* Varianzquelle manifestieren kann. Zum Beispiel kann sich in einem Generalfaktor bei Intelligenztests *Intelligenz plus Leistungsehrgeiz* ausdrücken, bei Fragebögen kann sich im Generalfaktor die mit den Items erfragte Traitdisposition zusammen mit einer mehr oder weniger ausgeprägten Akquieszenzneigung ausdrücken. Über die relativen Anteile der an einem Faktor beteiligten ‚Teil-Varianzquellen' ist aus den Daten selbst oder aus ihrer formalen Repräsentation keine Information zu gewinnen. Die Komplexität latenter Bedingtheiten kann sich mit Varimin nur in dem Maße differenziert ausdrücken, wie

---

71 In einem Projekt mit Reanalysen von NEOPI-R-Daten, das mithilfe der Varimin-Transformation von Faktoren zu einer neuen Konzeptualisierung von Persönlichkeitskomponenten führen sollte (Ertel b), wurden die Begriffe ‚Aktivationslevel', ‚Aktivationsgradient', ‚Regulationsquelle', ‚Phänomenqualität' und ‚Funktionsbilanz' zur Deutung der Varimin-Faktoren vorgeschlagen. Abstrakte Begriffe dieser Art liegen nicht immer vor, sie sind schwer zu bilden, doch sind sie als Komponenten psychischen Geschehens theoretisch leichter miteinander verknüpfbar.

sich die Varianzquellen mit *unterschiedlichen* Beiträgen an der Varianz der verwendeten Variablen beteiligen.

Dass sich die Interpretierbarkeit initialer komplexer Strukturen, wenn diese mehr als zwei Faktoren aufweisen, durch Varimin-Rotationen verbessert, haben einige Beispiele verdeutlicht. Mir ist bei zahlreichen faktoriellen Reanalysen kein Fall begegnet, bei dem das Ergebnis einer Varimin-Transformation *weniger* gut interpretierbar gewesen wäre als das Ergebnis der Initiallösung. [72]

Was die Interpretierbarkeit von SS-Strukturen betrifft, so ließ diese bei den ausgewählten Beispielen zu wünschen übrig. In der Praxis würde man bei manchen der hier verwendeten Datensätze Faktorenanalysen nicht einsetzen, man würde auf andere Verfahren, vornehmlich auf MDS (bzw. NMDS) oder MTMM oder Circumplex-Verfahren ausweichen. Der Vorzug des CS-modellierenden Verfahrens ist unter anderem darin zu sehen, dass es sich zur Analyse der verschiedensten Datensätze und für unterschiedliche Fragestellungen eignet, während das SS-modellierende Verfahren manche Anwendungen von vorne herein ausschließt.

Nur bei besonderen Fragestellungen kann eine FA mit SS-Rotation nützlich sein. Ihre Bedeutung besteht aber nicht in dem, was man bisher glaubte, dass mit ihnen ‚Dimensionen' der untersuchten Domäne aufgedeckt würden, sondern darin, dass die Variablen einer Domäne geclustert werden, was gegebenenfalls erwünscht sein kann. Allerdings bleiben dann die latenten cluster-bildenden Merkmale unaufgeklärt. Nur wenn eine Fragestellung zu bearbeiten sein sollte, die das Clustern von kovariierenden Variablen zum Ziel hat, ohne dass die Quellen ihrer Kovarianz interessieren, würden sich Varimax und die obliquen Rotationsverfahren – diese insbesondere – als hilfreich erweisen.

Varimax könnte darüber hinaus nützlich bleiben für den Fall, dass Variablen zur Analyse anstehen, von denen man vorweg weiß, dass ihnen nur *eine* Varianzquelle zugrunde liegt. Das wäre z. B. der Fall, wenn die mit Varimin aufgedeckten latenten Merkmale, etwa als Items für Fragebögen, also als manifeste Variablen verwendet werden und wenn eine Korrelationsmatrix dieser Variablen einer Faktorenanalyse übergeben wird.

Komplikationen können für Varimin-Untersuchungen auch dann entstehen, wenn a priori unbekannt ist, ob die zur Analyse ausgewählten Variablen tatsächlich zwei oder mehr Varianzquellen haben. Bislang bin ich davon ausgegangen, dass Variablen in aller Regel multifaktoriell bedingt sind. Dies aber schließt Ausnahmen nicht

[72] Die Gesamtzahl der Datensätze, an denen das Varimin-Programm erprobt wurde, ohne dass Ergebnisse publiziert wurden, liegt bei über 500.

aus, und man weiß nicht, wie viele Ausnahmen man zulassen darf und wie sie als solche zu erkennen sind.

Von Ungewissheit bei der Deutung Varimin-rotierter Faktoren war schon die Rede bei der Behandlung der Frage, ob und wie *negative* Faktorladungen zu deuten sind. Ob bei den mit Plus- und Minusvorzeichen bipolar gekennzeichneten Strukturen funktional antagonistische Verhältnisse vorliegen oder ob sich in den Vorzeichen lediglich das Vorhandensein bzw. Nichtvorhandensein einer latenten Bedingung ausdrückt, darüber sagt die Analyse selbst nichts aus. Darüber entscheidet man mit einschlägigem Kontextwissen und den daraus sich ergebenden Mutmaßungen.

Bei der Verwendung des neuen Analyseverfahrens ist zu beachten – wie zuvor angedeutet –, dass die Variablen-Stichprobe für die Domäne, deren Kovarianzquellen differenziert werden sollen, repräsentativ ist. Die Variablen sollten als Ausdruck einer Mehrzahl latenter Bedingungen verdächtig sein, die unter den Domänvariablen Kovariationen hervorrufen. Zumindest sollte man Variablen nicht aufnehmen, die plausiblerweise nur wenig Aufschluss über Kovariationsquellen liefern. Würde man z. B. altersabhängige Variablen unter die Variablen eines Persönlichkeitsfragebogens mischen („ich werde immer vergesslicher“, „mir geht es gesundheitlich nicht so gut wie früher“), dann würden diese möglicherweise einen eigenen Faktor („Altersanzeichen“) bilden. Diese würden zwar mit Plusladungen einen Beitrag zur Varianz bekunden – wenn ein solcher vorliegt – , sie würden aber bei den Variablen, die mit Alter nichts zu tun haben, vermutlich Minusladungen hervorbringen („Einfluss nicht vorhanden“ oder „hat mit Alter nichts zu tun“). Mit aufschlussarmen Minus-Ladungen bei diesen Variablen würde Kommunalität ‚verschwendet‘ werden. [73] Re-Analysen mit Varimin-Rotation führen mitunter durchaus auch zu unbefriedigenden Ergebnissen. Diese sind jedoch meist auf unzureichende Variablen-Stichproben zurückzuführen. Mithilfe eines Measures of Sampling Adequacy (Kaiser, 1970) ließe sich ggf. die Adäquatheit von Datensätzen für exploratorische Faktorenanalysen routinemäßig ermitteln.

Zur weiteren Agenda gehört die Klärung des Invarianzproblems. Die Frage bleibt zu prüfen, ob das Varimin-Verfahren zu Lösungen führt, die gegenüber Hinzufügungen zur Variablenstichprobe und gegenüber Streichungen aus dieser oder auch gegenüber einem Wechsel der untersuchten Personenstichproben stabil genug

[73] Die Faktorladung einer Varianzquelle, die bei einer Variablen nur ein ‚nicht vorhanden‘ aussagt, wird dennoch bei der betreffenden Variablen als Beitrag zur Kommunalität mit aufgenommen, was eine Art Informationsverschwendung ist, zumal wenn es sich um eine Varianzquelle handelt, die nur für eine Minorität von Variablen als ‚vorhanden‘ relevant wird.

sind. Eine nicht-publizierte Untersuchung (Ertel, b) ergab deutlich mehr Stabiliät für Varimin als für Varimax beim Hinzufügen einer artifiziellen Variablen zu einer zuvor analysierten Stichprobe naturalistischer Variablen.[74]

Zu den noch zu bearbeitenden Fragen gehört auch die nach der optimalen Faktoren-Extraktionsmethode. Im bisherigen Projekt wurde ausschließlich die Hauptkomponentenmethode (PCA) verwendet, weil sie sich in der faktorenanalytischen Forschung zum Standard entwickelt hat. Es mag sein, dass das Einsetzen von Einsen in die Diagonalfelder der Korrelationsmatrizen, das zur PCA gehört, auf die Endergebnisse u. U. eine leicht verzerrende Wirkung hat, vor allem dann, wenn überwiegend niedrige Interkorrelationen vorliegen. Zu den weiteren Aufgaben der Methodenforschung gehört, die Hauptachsen- und das Maximum-Likelihood-Verfahren vergleichsweise auf ihren Beitrag zur Modelliergüte zu prüfen und ggf. weitere Verbesserungen des Vorgehens vorzunehmen, falls durch eine andere Wahl der Extraktionsmethode am Ergebnis noch etwas zu verbessern sein sollte. [75]

Die vorliegende Arbeit beschränkte sich auf die Entwicklung von Varimin als einem Verfahren, das für die exploratorische Forschung eine neue Richtung zeigt. Nachdem sich das Verfahren in zahlreichen Prüfungen mit verschiedenen Datensätzen bewährt hat, wird sein Einsatz bei der systematischen Bearbeitung psychologischer Sachfragen zum Desiderat. Zwei abgeschlossene Projekte zur Frage der Varianzquellen der Intelligenz (s. Kapitel 3) und der Persönlichkeit (Ertel, c) liegen vor.

Abschließend sei ein allgemeinerer Einwand bedacht, der sich richten könnte gegen die dem vorliegenden Ansatz zugrunde liegende Strategie, über formale Fragen der Datenverarbeitung (hier: Faktorenrotation) zu Entscheidungen zu kommen aufgrund von Ergebnissen, die aus inhaltlichen Zusammenhängen gewonnen wurden. Mit Argumenten qualitativer Herkunft könne man ein mathematisches Instrumentarium nicht kritisieren. Diesem Einwand sollte man entgegenhalten, dass es in dieser Arbeit darum ging, einen Fehler der Methodiker zu korrigieren: die Dogmatisierung von SS als Modell für latente Bedingungen bei multi-

---

[74] Einen Mangel an Invarianz von SS-Lösungen beklagte schon Butler (1969): *"…the simple structure concept does not solve one of the most crucial and fundamental problems of factor analysis, the problem of the likelihood of factorial invariance"* (p.13). *"Normal varimax factors cannot be regarded as factorially invariant…"* (p. 24)

[75] Dies ist indessen nach bislang vorliegenden Vergleichen der Extraktionsverfahren kaum zu erwarten: *„The major conclusion of this article is that there is little basis to prefer component analysis or factor analysis. For practical purposes the choice of method is not a decision that will greatly affect empirical results or substantive conclusions"* (Velicer & Jackson, 1990, p. 19).

variaten Analysen. Dieses Modell ist ein formales, doch es wurde mit durchaus nicht-formalen Motiven eingeführt, denn die Behauptung seiner Notwendigkeit (SS-Transformation mache Faktorstrukturen erst interpretierbar) ergibt sich, wenn man die Güte der Methode auf die Inhalte der mit ihr gewonnenen Ergebnisse bezieht. [76] Dieser Bezug erfordert psychologisches *Verstehen*, die Mitverwendung von Wissen, das über statistisch belegbare Ergebnisse hinaus geht. Thurstonianer geraten mit ihrer Begründung des faktoriellen Rotationsprinzips, mit der sie dem nicht-formalisierten Verstehen Konzessionen einräumen, in die Nachbarschaft anderer Forscher, die sich auch – mehr noch als sie – dem Verstehen verpflichtet fühlen. Auf dem Forum, auf dem sich die Positionen der Modellformalisten und der weniger streng und eng denkenden Phänomenbeschreiber überschneiden, sollten Letztere ein Mitspracherecht geltend machen dürfen. Sie sollten auf mögliche Fehler der Formalisten achten dürfen, soweit sie deren Sprache verstehen, sollten die Fehler durchdenken, wenn sie nicht wegzudenken sind, und, soweit es in ihren Kompetenzbereich fällt, auch helfen dürfen, sie zu korrigieren.

[76] *„…consistent psychological meaning is by far the most important criterion for the success of a factor analysis that is designed to illuminate psychological phenomena. The simple structure criterion was designed only as a means to that end" (Guilford & Hoepfner, 1969, 6). "The arguments in favor of rotation are not mathematical; and in each research the investigator has to decide its merits on non-mathematical grounds"* (Burroughs & Miller, 1961, p. 35).

Kapitel 3

Komplexität intellektueller Leistungen.
Grundintelligenz (g) und Lernkapital (l).

Vorspann

In diesem Kapitel wird über eine Faktorenanalyse der Intelligenz berichtet. Intelligenz war die erste Forschungsdomäne des Pioniers faktorenanalytischer Methodik (Spearman, 1904, und seiner frühen Nachfahren (Cattell, Thurstone, Vernon, Burt). Inzwischen hat sich mit ‚simple structure' (SS)-orientierten Analysen zur Frage, welche Differenzierungen im Bereich der Intelligenz faktoriell zu unterscheiden sind, ein beachtlicher Konsens entwickelt. In den Lehrbüchern von heute unterscheidet man die von Cattell & Horn (1963) eingeführten Intelligenzfaktoren ‚fluide' Intelligenz (angeboren) und ‚kristalline' Intelligenz (durch schulisch-kulturelles Lernen erworbene Intelligenz). Mit dem Intelligenz-Struktur-Test (I-S-T), dem im deutschsprachigen Raum am häufigsten eingesetzten Intelligenztest, sollen die beiden Ausprägungsformen der Intelligenz erfasst werden.

Im dritten Kapitel dieser Berichtssammlung werden publizierte Daten aus I-S-T-Untersuchungen komplexstrukturell reanalysiert. Die Erwartung war naheliegend, dass die mit *Complex Structure Modeling* (CSM) gewonnenen Ergebnisse von denen abweichen, die mit *Simple Structure Modeling* (SSM) gewonnen wurden. Es kam darauf an, die Ergebnisse der beiden Analysemethoden vergleichend zu evaluieren.

Mich hat angenehm überrascht, dass die CSM der Daten eine klare Trennung nahe legte zwischen der Intelligenz als Grundbedingung intelligenter Leistungen (‚Grundintelligenz' *‚g'*, mit überwiegendem Anteil an Varianzaufklärung) und den hinzu kommenden Leistungsbedingungen, die dem schulisch-kulturellen Lerngewinn zuzuschreiben sind (‚Lernkapital', *l*, mit weit geringerem Anteil). Durch eine SSM von Intelligenztestdaten entstehen Mischprodukte, die diese beiden Quellen der Leistungsvarianz miteinander vermengen und zwei ungefähr gleich große Scheindimensionen hervorbringen.

Der Gewinn, der durch die komplexstrukturelle Analyse mit ihrer statistischen und begrifflichen Trennung der beiden Varianzquellen erzielt wird, wird bestätigt durch erhöhte Korrelationen der Varimin-Faktorscores (im Vergleich zu Varimax-Faktorscores) mit diversen testunabhängigen Leistungskriterien. Die vorliegende Untersuchung kann als methodisches Modell für weitere Untersuchungen zur Intelligenzdiagnostik gewertet werden, denn mit mehr Varianzquellen ist zu rechnen, denen voraussichtlich aber zunehmend geringere Aufklärungsanteile zukommen.

Zu den Vorteilen einer CSM von Intelligenztestdaten gehört: Auf den Begriff der Fähigkeit (ability oder capacity) darf und muss man verzichten. Dieser hat bislang das Standard-Suchschema der Intelligenzdiagnostiker abgesteckt und das Deuten von Faktoren eingeengt. Der Begriff der Fähigkeit wird hier durch den deutungsoffeneren Begriff der Varianzquelle ersetzt. Damit können auch verschiedene Ausprägungen des kognitiven Stils, der Interessen, der Leistungsmotivation als fördernde bzw. hemmende Bedingungen für Intelligenztestleistungen in Betracht kommen, die vordem in den Fähigkeitsbegriffen des SSM verschwanden.

Ausgangslage und Zielsetzung

Um der Komplexität latenter Determinanten gerecht zu werden, schien mir ein Gegenprinzip gegenüber SS gerade recht, welches Komplexstrukturen aufsucht, statt sie zu vermeiden. Dies habe ich mithilfe eines unkonventionellen Rotationsverfahrens (Varimin) zu realisieren versucht.[77] Varimin hat zum Ziel, die mit der Initiallösung schon vorliegende, aber oft noch defizitäre faktorielle Komplexität zu optimieren. Eine Variminrotation der Faktoren sucht das Gegenteil von dem zu erreichen, was eine Varimaxrotation zu erreichen sucht. Varimin minimiert die Varianz der Faktorvektoren, durch Varimax wird sie maximiert.

Um festzustellen, ob sich der Einsatz von Varimin bei faktoriellen Problemlösungen praktisch bewährt, wurden im zweiten Kapitel Ergebnisse von Varimin- und Varimaxtransformationen bei gleichen Datensätzen miteinander verglichen. Bei diesen und weiteren bislang noch unveröffentlichten Vergleichen(Ertel, a, b) ergaben sich regelmäßig Vorteile für die Faktoreninterpretation zugunsten von Varimin. Auch zeigte sich, dass – zumindest bei drei- und mehrfaktoriellen Analysen – die Varimin-Lösung auch im Vergleich mit der Initiallösung leichter interpretierbar ist, obgleich schon die Initiallösungen der Komplexität des jeweils untersuchten Sachverhalts in der Regel entgegen kommen. Wenn einem Datensatz nur zwei substantielle Faktoren zugrunde liegen, lässt sich die Initiallösung durch eine Varimin-Rotation kaum noch verbessern.

Im Folgenden soll die Leistung von Varimin an Daten des Intelligenztests I-S-T überprüft werden. Zur Deutung der Varimin-rotierten I-S-T Faktoren, die in Studie I geleistet werden soll, sollen in Studie II Korrelationen mit Außenkriterien beitragen. Dabei werden sich Vergleiche zwischen Varimin- und Varimaxlösungen als hilfreich erweisen.

[77] Für die statistische Programmierarbeit stand mir der Physiker Uwe Engeland zur Seite, der meine Anweisungen konkret umzusetzen verstand.

Studie I: Varimin-Analyse der I-S-T-Faktoren

Zielsetzung
Zunächst steht die Frage an, ob mit Einsatz der Variminrotation sinnvolle I-S-T-Faktoren resultieren.

Material
Um eine möglichst stabile Faktorenstruktur zu gewinnen, wurden 18 in der Literatur auffindbare Interkorrelationstabellen des I-S-T, jede mit 8 Subtestvariablen, faktorisiert. Die 18 Faktorladungen wurden für jeden Subtest gemittelt. Über die Quellen der Daten gibt Tabelle 3.01 Auskunft. Die Datensätze stammen aus Untersuchungen mit den I-S-T Versionen 1955 und 1970. [78]

Erwartung
Amthauer et al. (1999) hatten mit dem I-S-T, Testversion 2000, eine zweidimensionale faktorielle Struktur ermittelt, die die Autoren nach SSM-orientierter Rotation als Manifestation einer ‚fluiden' und ‚kristallinen' Intelligenz deuteten. Erwartet wird, dass die Faktoren des I-S-T nach einer Varimin-Rotation anders und befriedigender zu deuten sind als nach einer Standard-Varimax-Rotation.

Datenauswertung
Von den neun Subtests des I-S-T wurde ME (Merkfähigkeitstest) ausgeschieden, da auch die Rezensenten des I-S-T 70 (Schmidt-Atzert & Hommers, 1996, Schmidt-Atzert, 1997, Brocke et al. 1998) die Gedächtnisfunktion, die mit dem Subtest ME erfasst werden soll, durch diesen Test allein nicht hinreichend repräsentiert sahen. Die Mitverwendung von ME könnte sich auf die Faktorstruktur des Gesamttests störend auswirken. Auch wurden alle Nicht-I-S-T-Variablen aus geschieden (z. B. zur Persönlichkeit), wenn solche in den Original-I-S-T Datensätzen mit verwendet worden waren.

Die 18 Korrelationsmatrizen der verbleibenden 8 Subtestvariablen wurden pro Datensatz durch PCA faktorisiert, von jeder Auswertungseinheit wurden eine Zwei- und Dreifaktorenlösung nach Varimin und Varimax rotiert. Die gewonnenen Faktorgewichte wurden über die 18 Analysen hinweg mit Einschaltung von Fishers Z-Transformation gemittelt. Vor der Ermittlung dieser Durchschnittswerte wurden die Faktorenfolgen der 18 Analysen, soweit nötig, miteinander synchronisiert, denn nicht immer erschienen gleiche Faktoren in der Extraktionsfolge

[78] Von der revidierten Version 2000, die sonst bevorzugt worden wäre, lagen zuwenig Korrelationstabellen vor.

an gleicher Stelle. Auch wurden die Vorzeichen der Ladungen in den verschiedenen Analysen aufeinander abgestimmt. So kam es mitunter bei einem Faktor zur Vorzeichenumkehr, wenn diese erforderlich war, um eine gegensätzliche in eine

Tabelle 3.01

Datenmaterial. Die Untersuchungen 1-8 wurden mit dem I-S-T 55, die von 9 - 18 mit dem I-S-T 70 durchgeführt

Nr	Jahr	Autoren	Probanden	N	Faktoren	Zusätzliche Variablen
01	1958	Fischer, H.	Schüler (14-16 J.)	122	3 *	4 Variablen der Primary Mental Abilities
02	1964	Höger, D.	Gymnasialschüler (12-19 J.)	519	6 *	Noten von 11 Schulfachleistungen
03	1966	Bäumler, G. & Weiß, R.	Berufsschüler	200	6 *	10 Variablen: CFT Pauli-Test
04	1967	Haenschke, B. & Mehl, J.	Normalschüler (13.2 J.)	77	6 *	4 Tests Produktivität 5 Schulfachnoten
05	1967	Haenschke, B. & Mehl, J.	Spezialschüler (13.0 J.)	67	6 *	4 Tests Produktivität 5 Schulfachnoten
06	1969	Seitz, W. &. Löser, G.	Gymnasialschüler (17;6)	124	3 *	4 CFT, 7 Schulnoten 12 HSPQ (Persönlichkeit)
07	1970	Amthauer, R.	Stichprobe sehr heterogen (Vermutung)	799	X	Keine zusätzlichen Tests verwendet
08	1970	Bäumler, G. & Breitenbach, W.	Psychologiestudenten	55	7 *	Pauli-Test 5 Motivationsvariablen
09	1970	Sassenscheidt, H. & Buggle, F.	Handwerker, Techniker, Ingenieure	132	2	4 PTV-Variablen (techn. Verstdn) 16 PF
10	1972	Langner, E. & Olbrich, M.	Flugzeugführeranwärter	397	3 #	5 Variablen EVT (Intelligenztest)
11	1972	Langner, E. & Olbrich, M.	Abiturienten-Offiziersbewerber	190	3 #	5 Variablen EVT (Intelligenztest)
12	1972	Langner, E. & Olbrich, M.	Absolventen der mittleren Reife	210	3 #	5 Variablen EVT (Intelligenztest)
13	1972	Langner, E. & Olbrich, M.	Volksschulabsolventen	236	3 #	5 Variablen EVT (Intelligenztest)
14	1972	Langner, E. & Olbrich, M.	Wehrpflichtige	150	3 #	5 Variablen EVT (Intelligenztest)
15	1993	Schmidt-Atzert, L. et al.	Hauptschüler (17.5 J.)	196	2	CFT Diktat, Grundrechnen
16	1993	Schmidt-Atzert, L. et al.	Realschüler (18.1J.)	394	2	CFT. Diktat. Grundrechnen
17	1993	Schmidt-Atzert, L. et al.	Gymnasiasten (21.J.)	397	2	CFT. Diktat, Grundrechnen
18	1998	Brocke, B. et al.	Eignungstest-Probanden (38.8 J.)	279	2	FRT(Intelligentest)

Anmerkung zur Tabelle:
* Faktorenanalysen I-S-T nur zusammen mit zusätzlichen Variablen
Drei Faktoren werden als Bestlösung betrachtet, geprüft wurden 2–5 Faktorlösungen X Keine Faktorenanalyse durchgeführt

gleiche oder fast gleiche Struktur wie die eines Faktors der Mehrheit der anderen Analysen zu verwandeln. Dabei wurden die Ladungsmuster bei den verschiedenen Analysen durch visuelles Vergleichen abgeglichen. Ungewissheit bei dieser Entscheidung trat in zwei bis drei Fällen von Dreifaktorenlösungen auf. Da Unsicherheit nur auftreten konnte bei zwei zuzuordnenden Faktoren, deren Ladungsprofile sehr ähnlich waren, kann ein Effekt durch mögliche Fehlentscheidung, der minimal wäre, vernachlässigt werden.

Ergebnisse mit Erörterung

Die Faktorstruktur-Aggregate der Varimin- und Varimaxrotationen zeigt Tabelle 3.02. Zunächst interessieren die Unterschiede zwischen der Zweier- und Dreierlösung. In den 18 originalen I-S-T-Analysen, die mit SSM-Methoden vorgenommen worden waren, hatten die Autoren nicht selten einen dritten Faktor für varianzstark und interpretierbar gehalten. Man würde einen vom Aggregatdatensatz gewonnenen dritten Faktor akzeptieren, wenn er erstens einen hinreichenden Varianzanteil erkennen lässt und zweitens ein deutbares Ladungsprofil besitzt. Doch somit ist Varimin-F_3 zu ignorieren, nur Varimin-F_1 und F_2 sind als substantiell zu betrachten, im Varimin-Aggregat erreicht F_3 nur 2 Prozent der aufgeklärten Gesamtvarianz. [79]

[79] Nach bereits vorliegenden Erfahrungen mit Varimin, die noch spezifiziert werden müssen, sollte ein Faktor nicht weniger al 10% Varianz aufklären, um als Faktor mit einem substanziellen Beitrag zur Gesamtvarianz anerkannt zu werden.

Tabelle 3.02

Gemittelte Faktorladungen (über 18 Analysen) der Zweier-und Dreierlösungen nach Varimin- und Varimaxrotation der einzelnen Initiallösungen

	A. Varimin					B. Varimax				
Lösung:	Zweierlösg.		Dreierlösung			Zweierlösg.		Dreierlösung		
Deutung:	g	l	g	l		c	f	c	f	A
Subtest:	F_1	F_2	F_1	F_2	F_3	F_1	F_2	F_1	F_2	F_3
1 SE	.63	.37	.63.	35	-.12	.70	.24	.72	.16	.24
2 WA	.62	.32	.63	.33	.06	.67	.23	.65	.16	.43
3 AN	.70	.29	.68	.31	-.06	.69	.32	.59	.23	.42
4 GE	.65	.35	.63	.38	.17	.68	.31	.67	.18	.35
5 RA	.71	-.04	.69	-.22	-.19	.53	.53	.25	.45	.64
6 ZR	.68	-.21	.66	-.20	-.08	.39	.62	.25	.52	.50
7 FA	.65	-.38	.67	-.35	.11	.22	.72	.12	.72	.32
8 WU	.57	-.51	.59	-.38	.09	.11	.74	.10	.74	.29
h^2 Varianz %	**79**	**21**	**79**	**19**	**.02**	**54**	**46**	**38**	**34**	**28**

Anmerkung:
g = allgemeine oder Grund-Intelligenz l =Lernkapital
f = fluide c =kristalline Intelligenz A = Artefakt
SE Satzergänzungen, WA Wortauswahl , AN Analogien, GE Gemeinsamkeiten finden, RA Rechenaufgaben, ZR Zahlenreihen, FA Figurenauswahl, WU Würfelaufgaben

Beim *Varimax*-Faktorenaggregat fallen im Vergleich zu Varimin zunächst völlig andere Ladungsprozente auf. Bei der Zweierlösung ergeben sich für Varimax-F_1 vs. F_2 54% vs. 46% Ladungsanteile (gegenüber 79% vs. 21% bei Varimin-F_1 und -F_2). Bei der Dreierlösung ergeben sich für Varimax F_1, F_2, F_3 38%, 34%, 28% Ladungsanteile (gegenüber 79%, 19%, 2% bei den Varimin-Faktoren). Vor allem fällt der Kontrast der Ladungsprozente bei F_3 ins Auge (Varimax-F_3 28%, Varimin- F_3 2%). Den relativ hohen Ladungsanteil bei Varimax-F_3 hätte man zu begrüßen, wenn F_3 zu deuten wäre. Doch dies ist nicht der Fall, die Ladungen des Varimax-F_3 Aggregats sind bei den acht Subtests des I-S-T ungefähr gleich hoch, sagen also über unterschiedliche Testleistungsbedingungen nichts aus, die man

sonst bei ihnen vermuten könnte. Varimax-F_3 in Tabelle 3.02B erweist sich somit als Artefakt und darf im Weiteren ignoriert werden. [80]

Wie sind die Zweierlösungen bei Varimax und Varimin zu deuten?

Zu Varimax F_1 und F_2 (Tabelle 3.02B): Die F_1- und F_2-Profile der bisherigen Faktoranalysen zum I-S-T, die sich meist der Varimaxrotation bedienten, sind den Profilen des hier vorliegenden Varimax-Faktorenaggregats natürlich ähnlich. Amthauer et al. (2001) interpretieren nach einer Oblimin-Rotation, die wie Varimax SSM zugrundelegt (und dabei etwas Schiefwinkligkeit in Kauf nimmt), Oblimin-F_1 als Faktor der ‚kristallinen' Intelligenz, Oblimin-F_2 als Faktor der ‚fluiden' Intelligenz.[81] Diese Faktoren lassen sich nicht als Materialfaktoren deuten (nicht etwa z. B: F_1 *verbal*, F_2 *figural*), da die beiden Tests mit Zahlenmaterial (RA und ZR) sowohl Varimax-F_1- als auch Varimax- F_2-Ladungen haben.

Die Deutung der beiden Varimax-Faktoren als Manifestationen einer ‚fluiden' (genetisch ursprünglichen) und ‚kristallinen' (durch Lerneinflüsse erweitert und verfestigt) erschien früheren Intelligenzforschern plausibel: Intelligenztest-Leistungen sind je nach Testart auch abhängig vom vorausgehendem schulisch und kulturell geförderten Lernen. Cattell (1963), Horn (1976) und andere Autoren sind mit SSM-orientierten Faktorenanalysen auf schulisch-kulturelle Einflüsse schon oft gestoßen, wenn durch Faktoranalysen zweiter Ordnung eine höhere Ebene der konzeptuellen ‚Hierarchie' erreicht wurde. Selbst diejenigen Forscher, die mit herkömmlichen Analyseverfahren eine kristalline Intelligenz in ihren Daten nicht fanden (Johnson & Bouchard, 2005), gehen davon aus, dass das der Testleistung vorausgehende Lernen zur Leistungsvarianz merklich beitragen kann (p. 410). Von einer Faktorenanalyse sollte man erwarten, dass sie die Varianzquelle des Lernens, die die Varianzquelle der genetischen Anlage ergänzt, erkennen lässt.

Doch wird dieser Erwartung durch SSM tatsächlich nicht entsprochen. Durch Varimax verliert die Hauptvarianzquelle der Intelligenz ihre Einheit, die sich

---

[80] Da SSM-Rotationen die initialen *g*-Faktorgewichte auf die nachfolgend extrahierten Faktoren gleich welchen Inhalts großenteils abgeben (s. unten), ist denkbar, dass es sich beim vorliegenden Varimax-F_3–Faktor um aggregierte Restbestände von *g* handelt, die schon in den 18 diversen I-S-T-Datensätzen auf dort extrahierte Artefakt-Faktoren verteilt vorkamen, wegen dort nicht vorhandener inhaltlicher Bedeutung in den verschiedenen Untersuchungen in eher zufallsbedingt-unsystematischer Weise.

[81] Oblimin- und Varimax-Ergebnisse unterscheiden sich nur geringfügig.

schon mit F_1 der Initiallösung als eine Annäherung an *g* manifestiert, d. h. durch eine Varimax-Rotation wird F_1-Varianz zu einem beträchtlichen Teil an F_2 abgegeben. Damit werden zwei verschiedene Intelligenzarten geschaffen, denen die Einheit *g* geopfert wird. Der Anteil des *Lerneinflusses* am Zustandekommen der Leistungen, der sich tatsächlich nur leicht leistungsmodifizierend auswirkt, wird durch Vermischung mit *g*-Anteilen aufgebläht und verfälscht. Das SS-Rotationsergebnis beschreibt Jensen (1998) wie folgt: „*… So if you ask where g went, the answer is that it has been divided up and lies 'hidden' among all of the tests' smaller loadings on all of the orthogonally rotated factors. Its variance has not disappeared, it has simply been obscured by being dispersed throughout the whole factor matrix*" (Jensen, 1998, p. 66).

Zu Varimin- F_1: Mit Varimin-F_1 manifestiert sich die relativ konstant zu denkende Leistungsbedingung (*g*), die das Gesamtniveau der Testleistungen bestimmt. Der Ausdruck *Grundintelligenz g* hebt die Bedeutung von Varimin-F_1 als Basisbedingung hervor, die abgesehen von Zusatzbedingungen der Testleistung zugrunde liegt. Diese Deutung wird in Studie 2 durch Korrelationen zwischen Varimin-F_1 und der Gesamt-I-S-T 70-Leistung sowie durch Korrelationen zwischen Varimin-F_1 mit kulturfreien Tests geprüft (s. u.). [82]

Zu Varimin- F_2: Varimin-F_2 ist ein bipolarer Faktor. Negative Ladungsvorzeichen werden traditionellerweise bei Faktoren der Intelligenz nicht geduldet und durch die SS-Rotation beseitigt.[83] Die beim Einsatz der CS-Rotation auch bei Intelligenztest-Daten zugelassene Bipolarität wurde in Ertel, 2009b, ausführlich begründet. Mit Varimin-F_2 beim I-S-T manifestiert sich – zunächst hypothetisch – die lediglich leistungsmodifizierende Zusatzbedingung des schulisch-kulturellen Lernens. Varimin- F_2 ist bei den Subtests mit Wörtern und Sätzen (SE, WA, AN, GE) relativ stark positiv geladen, worin sich der Einfluss der hauptsächlich verbal-symbolisch trainierenden Erziehung ausdrücken kann. Die F_2–Ladungen bei den figuralen Aufgaben FA und WÜ haben ein negatives Vorzeichen. Dies lässt sich als relatives Defizit von Lernvorteilen bei FA und WÜ deuten, nicht etwa als Einfluss mit leistungsmindernder Wirkung. Beim schulischen Training werden sprachfreie formale Operationen an visuellen Formen selten gefordert. Für die Aufgaben mit Zahlen, bei denen Varimin-F_2-Ladungen um Null liegen, darf man einen geringeren schulischen Einfluss annehmen als für sprachliche Aufgaben, jedoch einen größeren als den für den Umgang mit abstrakten Figuren.

---

[82] Die Korrelationen zwischen Varimax F_1 und F_2 mit dem Standard-Gesamtwert des I-S-T 70 betragen .74 bzw. .63.

[83] Berneyer (1957) referiert in diesem Sinne: „*The different methods of [factor] analysis [of mental aptitudes] yield factors which have negative loadings … Such factors, so Thurstone contends, must be devoid of 'scientific meaning'. They do not permit us to 'interpret the various tests as functions of the mental aptitudes which those tests elicit'*" (p. 23).

Varimin-F_2 benötigt einen Namen. Ich schlage die Bezeichnung ‚Lernkapital' vor, sie lässt metaphorisch anklingen, dass sich Lerneffekte durch Speicherung anhäufen und bei Intelligenz-erfordernden Operationen ähnlich wie Zinsen eines Kapitals positiv auszahlen.[84] Bei den Tests mit positiver F_2-Ladung wirkt sich Lernkapital stärker aus als bei den Tests mit negativer F_2-Ladung. Es wäre falsch, die negativen Ladungsrichtungen bei F_2 als eine Beeinträchtigung der Testleistungen durch kulturell-schulisches Lernen auszulegen. Nur in seltenen Ausnahmefällen sind Testleistungen denkbar, die durch vorheriges Training benachteiligt werden. Man darf stattdessen davon ausgehen, dass manche Testleistungen durch Training weniger profitieren als andere, dies allein kann zur Polarität der Ladungsvorzeichen bei F_2 führen. Darüber geben die Testleistungen selbst keinen Aufschluss, nur mithilfe des Kontextwissens lässt sich darüber entscheiden.

Die hier vorgenommene Differenzierung entspricht dem ‚Dreieck' *Begabung, Wissen und Lernen* (Weinert, 1996). Zur *Begabung* gehört nach Weinert in erster Linie Grundintelligenz, das *Wissen* entspricht dem angehäuften Lernkapital. Das *Lernen* als dritte Dreiecksseite wird in der vorliegenden Untersuchung nicht auch erfasst. Es handelt sich um den Vorgang, der im Längsschnitt des zurück liegenden Lebens den Zuwachs an ‚Wissen' und ‚Können' hervorgebracht hat (Waldmann et.al, 2003). Untersuchungen zur Validität von Varimin-F_2 folgen in Studie II.

Mit einer Deutung von Varimin-F_2 als Lernkapital soll nicht das letzte Wort gesprochen sein. Denkbar wäre, dass mit F_2 die Präferenz einer von zwei polaren Denkstilen zum Ausdruck kommt. In den vornehmlich sprachlichen Subtests (F_2-positiv) könnten sich ganzheitlichere kognitive Operationen auswirken, in den figuralen Subtests (F_2-negativ) eher analytisch-zergliedernde Operationen. Bei den numerischen Tests könnten sich die beiden Operationen die Waage halten. Es könnte auch sein, dass schulisches ‚Lernkapital' leichter von Schülern mit ganzheitlichem Denkstil erworben wird und dass dem ganzheitlichen Denkstil auch genetische Dispositionen zugrunde liegen. Eine endgültige Entscheidung darüber zu treffen, ist hier nicht möglich und nicht nötig.

In jedem Fall ist von der Loslösung des Lerneinflusses (F_2) von der allgemeinen Intelligenz (‚g', F_1) ein theoretischer und empirischer Gewinn zu erwarten. Die Kontroversen, die durch Korrelationen zwischen dem seinerzeit undifferenziert ermittelten ‚g' und der Zugehörigkeit der Probanden zur weißen oder schwarzen

[84] Die ebenfalls dem Bereich der Ökonomie entnommene sehr ähnliche Metapher des ‚Investierens' stammt von Cattell (1971), der mit seiner „investment theory" einen Zusammenhang zwischen der ‚fluiden' und ‚kristallinen' Intelligenz herzustellen versuchte: Die ‚fluide' Intelligenz wird nach Cattell über die gesamte Lebensspanne durch Lernen 'investiert'(Holling et a., p. 21).

Rasse ausgelöst wurden und erbittert verliefen, wären bei Mitberücksichtigung des Lernkapitals (Varimin-F_2), das im Unterschied zu Varimin-F_1 durch Training und kulturelle Bedingungen gesteigert werden kann, mit Sicherheit glimpflicher verlaufen.

Zusammenfassung

Bei einer Dreifaktorenlösung zeigen die gemittelten Faktorladungen der Subtests des I-S-T einen zu vernachlässigenden dritten Faktor, wenn sie mit Varimin rotiert werden. Bei den gemittelten Varimaxfaktoren zieht zwar ein dritter Faktor beträchtliche Ladung auf sich, doch über die 18 Analysen hinweg mit unsystematisch wechselnden Profilen. Varimax-F_3 wird als Artefakt ignoriert.

Die konventionelle Deutung von F_1 und F_2 des I-S-T, die in Varimax-F_1 eine ‚kristalline' und in Varimax-F_2 eine ‚fluide' Intelligenz repräsentiert sieht, wird für die faktorielle Varimin-Lösung verworfen. Varimin-F_1 wird aufgefasst als Repräsentanz des Generalfaktors der Intelligenz (*g*), Varimin-F_2 als eine Zusatzquelle der Leistungsvarianz, als das von der Intelligenz relativ unabhängig aufzufassende Lernkapital (*l*), von dem die Probanden je nach vorausgehendem Lernaufwand/Lernangebot und je nach Subtest unterschiedlich profitieren.

Studie II: Zur Validierung der Varimin-rotierten I-S-T-Faktoren

Untersuchungsziel

Die Validität der Varimin-Faktoren F_1 und F_2 , die als *g* (Grundintelligenz) und *l*, (Lernkapital) gedeutet wurden, soll im Folgenden überprüft werden. Zusätzlich werden Vergleiche mit den Varimax-Faktoren der ‚fluiden' und ‚kristallinen' Intelligenz durchgeführt. Hierzu werden neben den I-S-T Daten herangezogen: 1. Schulische Gesamtleistungen aus Untersuchungen von Höger (1964) und Cronemeyer (1983); 2. Rechtschreib- und Rechenleistungen sowie Leistungen aus einem kulturfreien Intelligenztest aus einer Untersuchung von Schmidt-Atzert et al. (1995); und 3. Daten von zwei kulturfreien Intelligenztests aus einer Untersuchung von Brocke et al (1998).

1. Korrelationen mit der Schulleistung (Daten: Höger und Cronemeyer)

Material
Unabhängig voneinander haben Höger (1964, p.435) und Cronemeyer (1983, p. 172) neben I-S-T Testdaten die Schulnoten der Teilnehmer berücksichtigt (vgl. Tabelle 303). Die Schulnoten Högers stammen von 519 Gymnasialschülern der vier Oberstufen, die Cronemeyers von 656 Abiturienten.

Tabelle 3.03

Korrelationen zwischen F_2 (Varimin) der I-S-T Subtests (Zeile 1) mit Schulnoten von Höger (2) und Cronemeyer (3).

		SE	WA	AN	GE	RA	ZR	FA	WU	r	
1	$F_{2\ Varimin}$	*.37*	*.32*	*.29*	*.35*	*-.04*	*-.21*	*-.38*	*-.51*	zwischen F_2 und	*p*
2	$r_{Höger}$	.09	.12	.25	.17	.12	.09	.08	.06	$r_{Höger}$: .62	.05
3	r_{Cronem}	.17	.21	.32	.25	.30	.23	.08	.07	r_{Cronem}: .63	.05
4	r_{Mittel}	.14	.17	.29	.22	.22	.17	.08	.07	r_{Mittel}: .67	.03

Anmerkung: SE Satzergänzungen, WA Wortauswahl , AN Analogien, GE Gemeinsamkeiten finden, RA Rechenaufgaben, ZR Zahlenreihen, FA Figurenauswahl, WU Würfelaufgaben

Erwartung
Wenn Varimin F_2 als Lernkapital richtig gedeutet wurde, sollte die F_2-Varianz der Subtests des I-S-T den Zusammenhang mit der Schulleistung widerspiegeln.

Auswertung mit Erörterung
Die Varimin-Faktorgewichte der acht Subtests, gewonnen aus dem Faktorenaggregat der Untersuchung (Teil I, eigenes Ergebnis), und die Korrelationen zwischen den acht I-S-T Subtests mit Schulzeugnisnoten (Ergebnisse von Höger und Cronemeyer) stehen zur Auswertung an. Ablesebeispiel in Tabelle 3.03: Die Subtest-SE-Punktzahlen der Höger-Schüler korrelieren zu .09 mit dem Schulnotendurchschnitt dieser Schüler (Zeile 2). Die acht Korrelationskoeffizienten der Zeilen 2 und 3 (die Zeilen selbst korrelieren miteinander zu r = .77) und deren Mittelwerte in Zeile 4 wurden sodann mit den in Zeile 1 wiedergegebenen Varimin-F_2- Faktorwerten (s. Tabelle 3.02) korreliert. Die drei Korrelationen in der vorletzten Spalte von Tabelle 3.03 sind signifikant. Sie besagen, dass die Varianz der schulischen Lernleistungen mit der Varianz der F_2 -Ladungen der I-S-T-Subtests zusammenhängt. Die Deutung von F_2 als Lernkapital sieht sich gestützt.

Varimin-F_2 (Lernkapital) ist von Varimin-F_1 (Grundintelligenz) faktoriell unabhängig. Wohl aber sollte die Schulleistung auch von der Intelligenz der Schüler abhängen. Verwendet man Varimin-F_1 zusammen mit F_2 für eine multiple Regression mit Schulleistung (Zeile 4) als abhängige Variable, dann erhöht sich die Korrelation von .67 (Schulleistungsbezug zu F_2 allein) auf $r = .80$. Da die Originaldaten der Schüler von Höger und Cronemeyer nicht vorliegen, kann die Berechnung einer Korrelation zwischen den Schulleistungen und den *Faktorscores* F_1 und F_2 auf der Personenebene, die hier anstelle der Faktorladungen der Subtests zu bevorzugen wäre, nicht durchgeführt werden.

2. Korrelationen mit einem kulturfreien Intelligenztest, mit Leistungen im Rechtschreiben und Grundrechnen (Daten: Schmidt-Atzert).

Material
Eine von Lothar Schmidt-Atzert zur Verfügung gestellte Datenbank (in Teil I schon mit verwendet, Tabelle 3.01, Nr. 15-17) mit anonymisierten Testergebnissen von 980 Personen enthielt neben Originaldaten des I-S-T 70 die individuellen IQ-Werte, die mit dem sprachfreien Culture-Fair-Intelligence Test (CFT) gewonnen wurden (CFT-IQ), sowie Ergebnisse eines Rechtschreibdiktats und einer schultypischen Prüfung im Grundrechnen. Die getestete Stichprobe enthielt 397 Gymnasiasten, 394 Realschüler und 196 Hauptschüler. Die im Folgenden verwendeten Personenzahlen haben wegen nicht kompensierbarer Lücken bei einzelnen Variablen (missing data) meist ein etwas geringeres N.

Auswertung mit Erörterung
Den Auswertungen wurden neben den Varimin- F_2-Faktorscores, die als Maße des Lernkapitals interpretiert werden, auch nicht-faktorielle Lernkapital-Maße zugrunde gelegt, die aus den I-S-T Ergebnissen gewonnen wurden. Hierdurch sollte festgestellt werden, ob mit I-S-T-Originaldaten – ohne Verwendung von Faktorenanalyse und abgeleiteten individuellen Faktorscores – Schätzwerte für das Lernkapital zu gewinnen sind.

Ermittlung der Schätzwerte L_p und L_D für Lernkapital.
Die I-S-T Subtests SE, GE repräsentieren den positiven Pol von Varimin-F_2, die Subtests FA, WU den negativen Pol. Alle Subtestvariablen wurden in Standardwerte transformiert (Schmidt-Atzert hatte dies schon besorgt, Mittelwert 100). Sodann wurden die Lernindikatoren L_p und L_D gebildet (L = Lernindikator, p = Proportion, D = Differenz:

$L_{p} =$ (SE+GE)/(SE+GE+FA+WU)*100 (Mittelwert = 50)
L_{D} =(SE+GE)-(FA+WU) (Mittelwert = 0)

Erwartungen
(1). Die Faktorscores von Varimin-F_1 (Grundintelligenz) sollten mit den CFT-IQ-Werten hoch korrelieren. Varimin-F_2 (Lernkapital) sollte mit dem CFT-IQ nicht merklich korrelieren.
(2). Die Faktorscores von Varimin-F_2 (Lernkapital) sollten mit den beiden Schätzwerten des Lernkapitals (L_p und L_D) hoch korrelieren.
(3). Die Rechtschreibtest-Ergebnisse sollten nicht nur hoch mit den Varimin-F_1-Faktorescores (Grundintelligenz), sondern auch mäßig hoch mit den Faktorscores von Varimin-F_2 (Lernkapital) korrelieren.
(4). Die Leistungen im Grundrechnen sollten nicht nur hoch mit den Varimin-F_1-Faktorescores (Grundintelligenz), sondern auch mäßig hoch mit den Faktorscores von Varimin-F_2 (Lernkapital) korrelieren.

Die Ergebnisse zu den Erwartungen (1) und (2) enthält Tabelle 3.04.

Tabelle 3.04

Interkorrelation der kritischen Variablen. Hauptvariable Prozent-Lernkapital.

		Indikatoren für „Lernkapital“ (L)			g -Indikatoren	
		1 % Lern-kapital (L_P) *	2 Differenz (L_D) **	3 Varimin-F_2	4 Varimin-F_1	5 CFT-IQ
1	% Lernkapital (L_P)	1	.94	.93	-.03	.09
2	Differenz (L_D)	.94	1	.99	-.03	.09
3	Varimin-F_2	.93	.99	1	.00	-.10
4	Varimin-F_1	-.03	-.03	.00	1	.73
5	CFT-IQ	.09	.09	-.10	.73	1

Die Korrelationen lassen für das weitere Vorgehen günstige Voraussetzungen erkennen:

Zur Erwartung (1): Mit Varimin-F_1 (Grundintelligenz) korreliert der CFT-IQ hochsignifikant (.73), mit Varimin-F_2 (Lernkapital) liegt die Korrelation bei -.10, beides im Sinne der Erwartung, womit die Deutungen von Varimin-F_1 und -F_2 bestätigt werden.

Zur Erwartung (2): Die Faktorscores von Varimin- F_2 (Lernkapital) korrelieren mit L_P .93 und mit L_D .99. Die Höhe dieser Korrelationen ist insofern bemerkenswert, als den Faktorscores F_2 die Analyse aller acht I-S-T-Subtests zugrunde liegen, während die Indikatoren L_P und L_D sich nur auf vier Subtestergebnisse stützen. Auch ist zufrieden stellend, dass die beiden L-Indikatoren nicht mit den Faktorscores von Varimin-F_1 und nicht mit dem CFT-IQ korrelieren (beide Korrelationen betragen .09).[85] Somit wird man in zukünftigen I-S-T Untersuchungen das Lernkapital individueller Testteilnehmer durch Verwendung der hier sich bewährenden Differenz- (L_D) oder Quotientenwerte (L_P) abschätzen dürfen.

Zur Erwartung (3). Im Folgenden wird die Rechtschreibleistung der Probanden in differenzierter Weise mit L_P korrelativ in Beziehung gesetzt, was anhand von Abbildung 3.01 zu erläutern ist. Abbildung 3.01 zeigt, wie sich die verwertbaren Diktatleistungen von 945 Personen der Schmidt-Atzert-Stichprobe über die L_P- Skala verteilen. Die beiden Variablen, L_P und Rechtschreibdiktat, korrelieren mit $r = .30$ (die L_D-Variable, nicht abgebildet, korreliert ähnlich hoch, $r = .29$). Unabhängig von der Intelligenz trägt also das Lernkapital zur Diktatleistung einiges bei.

[85] Zwar darf man davon ausgehen, dass intelligentere Personen mehr lernen und ihre Erfahrungen besser auswerten als weniger intelligente, doch geht dieser Leistungsunterschied in die L-Indikatoren (Prozent- oder Differenzwert sowie F_2) nicht mit ein. Die L-Indikatoren sind wie F_2 auf das jeweils individuelle Maß der allgemeinen Intelligenz hin relativiert.

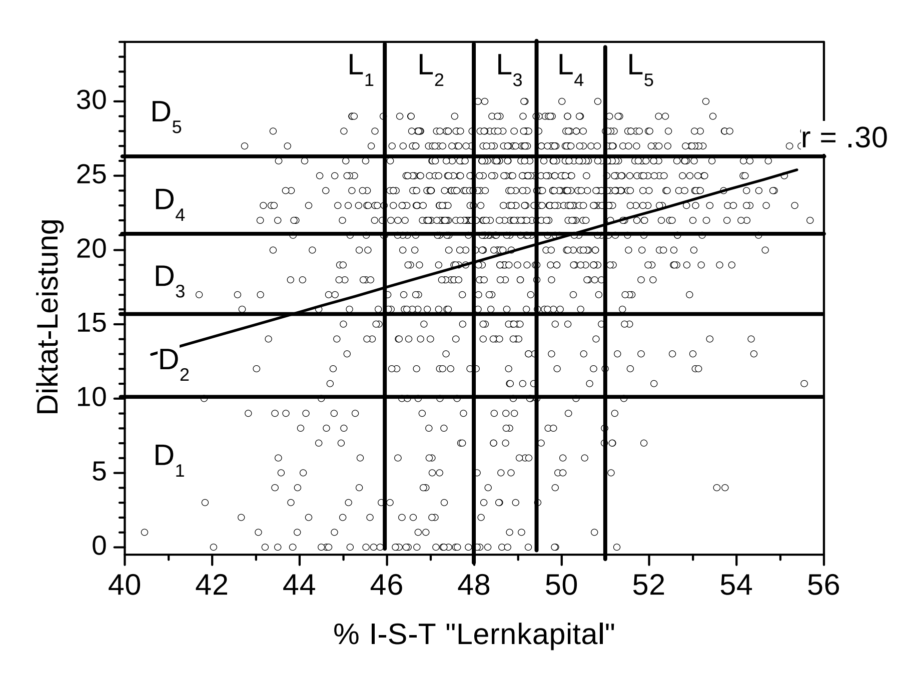


Abbildung 3.01

Verteilung der Diktatleistungen (Ordinate) über die Skala
des Lernkapitals (Abszisse). Differenziert wird
nach fünf Leistungsgruppen (D1 bis D5)
und fünf Gruppen Lernkapital (L1 bis L5).

Die Intelligenztestwerte (Varimin F_1-Faktorscores) sind indessen mit der Diktatleistung höher korreliert ($r = .60$), deshalb befinden sich die Personen mit weniger Intelligenz bevorzugt am unteren Ende der Y-Skala ‚Diktat', diejenigen mit höherer Intelligenz weiter oben (die Varianz der Intelligenz wird hier nicht auch als Koordinate dargestellt).

Wenn man die Gesamtstichprobe von unten nach oben durch horizontale Schnitte schrittweise in fünf Gruppen zunehmender Diktatleistung teilt (D_1 bis D_5) und diese Gruppen nochmals von links nach rechts in fünf Untergruppen zunehmenden Lernkapitals (L_1 bis L_5), dann wird, was die mittlere Intelligenz in den Untergruppen betrifft, Folgendes erwartet. Die mittlere Intelligenz (F_1-Faktorscores) nimmt von D_1 bis D_5 (von unten nach oben) mit der Diktatleistung zu, sie nimmt

in den Gruppen L_1 bis L_5 (von links nach rechts) mit zunehmendem Lernkapital ab. Warum? Je mehr Lernkapital die Probanden zum Rechtschreibdiktat mitbringen, umso weniger intelligent müssen sie sein, um ein bestimmtes Leistungsniveau zu erreichen. Man kann dasselbe auch anders ausdrücken: Lernen kompensiert bis zu einem gewissen Grade Begabungsmängel.

Tabelle 3.05

Mittlere Faktorscores F_1 (*g* -Indikator mit Standardabweichung SD), differenziert nach Lernkapital L_P (5 Stufen) und Diktatleistung (5 Stufen). Die horizontale und vertikale Auflistung der Gruppen L_P 1 bis 5 und der Diktatleistungen 1 bis 5 ist hier gegenüber der Anordnung in der zugehörigen Abbildung 3.01 vertauscht. N = 945.

Lern-kapital L_P	Diktat 1			Diktat 2			Diktat 3			Diktat 4			*Diktat 5*		
	N	**F_1**	SD	N	**F_1**	SD	N	**F_1**	SD	N	**F_1**	SD	*N*	***F_1***	*SD*
1 0-46	32	**-0.94**	.46	11	**-0.41**	.42	22	**-0.29**	.70	35	**0.43**	.63	*14*	***0.78***	*.74*
2 46-48	26	**-0.99**	.56	16	**-0.55**	.60	35	**-0.21**	.81	92	**0.52**	.98	*46*	***1.02***	*.89*
3 48-50	27	**-1.15**	.59	23	-**0.77**	.53	40	**-0.33**	.73	76	**0.47**	.77	*53*	***0.72***	*.82*
4 50-51	13	**-1.22**	.74	10	**-1.09**	.55	34	**-0.11**	.91	69	**0.32**	.72	*54*	***0.50***	*.88*
5 51-60	08	**-1.56**	.41	14	-**0.95**	.87	26	**-0.31**	.85	94	**-0.01**	.85	*75*	***0.16***	*.81*

Die Ergebnisse in Tabelle 3.05 entsprechen dieser Erwartung: In vier von fünf Diktatleistungsgruppen sinkt F_1 (Intelligenz) mit zunehmendem Lernkapital L_p,(von Zeile 1 oben bis zur Zeile 5). Nur bei der mittleren Diktatleistungsgruppe Nr. 3 ist kein Änderungstrend erkennbar.

Das Zusammenwirken von Intelligenz und Lernkapital bei der Diktatleistung hätte man auch pauschal, aber mit weniger Transparenz, durch eine multiple Regression ermitteln können (als Prädiktoren Varimin-F_1 und F_2 , als AV die Diktatleistung). Die multiple Korrelation, die F_2 einschließt, erhöht die Korrelation von r = .60 (Diktat mit Intelligenz F_1 allein) auf r = .69 (Diktat mit Intelligenz F_1 plus Lernkapital L_p), der Unterschied ist sehr signifikant (Z = 3.36, p = .0004). [86]

[86] Eine multiple Regression auf die Diktatleistung durch *Varimax* F_1 und F_2 kommt zum genau gleichen Ergebnis wie die multiple Regression mit *Varimin* F_1 und F_2. Denn die Gesamtinformation von Varimin F_1 und F_2 ist in der Gesamtinformation von Varimax F_1 und F_2 enthalten, nur ist sie dort auf F_1 und F_2 anders verteilt worden.

Die Zusammenhänge zwischen dem I-S-T und den Variablen CFT-IQ, Rechtschreiben und Grundrechnen werden, nach Schultypen differenziert, in Tabelle 3.06 wiedergegeben, sowohl für die Varimin- als auch diesmal vergleichsweise für die Varimax-Faktoren. Die Ergebnisse für Varimax folgen später.

Tabelle 3.06A (nochmals zur Erwartung 1): Der CFT-IQ (Grundintelligenz) korreliert sehr hoch mit Varimin- F_1 (Grundintelligenz) (.73), und nur geringfügig mit Varimin-F_2 (.11), also mit dem Faktor, auf dem diejenigen Subtests des I-S-T positiv laden, bei denen das schulisch-kulturelle Lernen Leistungsvorteile mit sich bringt.

Tabelle 3.06B: (nochmals zur Erwartung 3): Die Leistungen im Rechtschreibdiktat, die durch schulisches Lernen gefördert werden, korrelieren bedeutend höher als der CFT-IQ mit den Faktorscores von Varimin-F_2 , Lernkapital (.31). Die Korrelation zwischen Rechtschreiben und den Faktorscores von Varimin-F_1 (Grundintelligenz) fällt dementsprechend vergleichsweise niedriger aus als beim CFT.IQ (.60).

Zur Erwartung 4 (Tabelle 3.06C): Das Ergebnis für das Grundrechnen entspricht der Erwartung weniger durchschlagend. Neben der erwarteten hohen Korrelation mit den Faktorscores von Varimin-F_1 (Grundintelligenz) (.64) liegt eine nur eben signifikante ($p = .05$) Korrelation mit den Faktorscorers von Varimin-F_2 (Lernkapital) vor (.07).

Die Korrelationen mit den beiden Varimaxfaktoren sind weniger überzeugend. Die Faktorscores der ‚fluiden' Intelligenz (Varimax-F_2) hätten mit dem CFT-IQ bedeutend höher korrelieren sollen als die Faktorscores der ‚kristallinen' Intelligenz (Varimax-F_1). Doch sie korrelieren niedriger (.44 ‚fluide' I. vs. 59 ‚kristalline' I.). Nur korrelieren die Leistungen beim Rechtschreibdiktat und beim Grundrechnen erwartungsgemäß höher mit der ‚kristallinen' Intelligenz als mit der ‚fluiden' Intelligenz (Grundrechnen .64 ‚kristalline' I. vs. 21 ‚fluide' I., Grundrechnen .50 ‚kristalline' I. vs. 40 ‚fluide' I.). Die Vergleiche der Rotationsvaliditäten Varimin vs. Varimax fallen demnach erneut zugunsten von Varimin aus.

Tabelle 3.06

Pearson-Korrelationen zwischen den Varimin- und Varimax-Faktorscores F_1 und F_2 aus dem I-S-T und dem CFT-IQ, einem Diktat und einer Prüfung im Grundrechnen.

		A. CFT-IQ				B. Rechtschreibdiktat				C. Grundrechnen			
Rotation:		Varimin		Varimax		Varimin		Varimax		Varimin		Varimax	
Faktordeutung:		g	L	c	f	g	L	c	f	g	L	c	f
Schultypen:	N	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
Gymnasium	381	.66	-.10	.40	.53	.44	.18	.44	.18	.55	-.14	.29	.48
Realschule	198	.60	.12	.50	.34	.46	.26	.51	.13	.42	-.01	.29	.31
Hauptschule	381	.60	.02	.44	.41	.47	.27	.52	.14	.39	-.05	.24	.31
Gesamt	970	.73	.11	.59	.44	.60	.31	.64	.21	.64	.07	.50	.40

Anmerkung:
g = allgemeine Intelligenz L = Lernkapital
f = fluide c = kristalline Intelligenz

3. Korrelationen mit einem kulturfreien Intelligenztest (Daten: Brocke).

Material
Brocke et al. (1998) druckten in ihrem Artikel eine Interkorrelationstabelle ab (hier schon verwendet als Beitrag zu Teil I, s. Tabelle 3.01, Zeile 18), bei der sie neben den acht I-S-T Subtests die Korrelationen mit dem Figure-Reasoning-Test (FRT) einschlossen (N = 241 für FRT, N = 279 für I-S-T). Die Autoren erhofften sich vom Einbezug des FRT, der mit dem ebenfalls figuralen SPM von Raven .93 korreliert, „Anhaltspunkte für die interne Validität des I-S-T 70“ (p. 94), insbesondere Korrelationen mit den ‚fluiden‘ Subtests des I-S-T 70.

Erwartung
Nach einer Faktorenanalyse der Interkorrelationen sollte der FRT beim Variminfaktor F_1 hoch laden, sofern Varimin-F_1 als allgemeiner Intelligenzfaktor richtig gedeutet wurde. Der FRT sollte beim Variminfaktor F_2 eine niedrige Ladung zeigen, sofern Varimin F_2 als Lernkapital-Faktor richtig gedeutet wurde. Die Varimaxfaktoren F_1und F_2 , denen ‚fluide‘ und ‚kristalline‘ Intelligenz nachgesagt wird, sollten diesen Unterschied hinsichtlich der faktoriellen Validität nicht oder weniger deutlich zeigen.

Auswertung und Erörterung
Die Ergebnisse einer Varimin- und Varimaxanalyse der acht I-S-T-Variablen (Subtest ME wieder ausgeschlossen) denen die FRT-Variable hinzugefügt wurde, gibt Tabelle 3.07 wieder. Es zeigt sich:

Für Varimin: Der FRT hat eine zu erwartende hohe Varimin-F_1-Ladung (Grundintelligenz, .76) und eine ebenfalls zu erwartende niedrige F_2 -Ladung (-.15), denn Vorteile durch schulisch-kulturelles Lernen fehlen beim FRT ebenso wie bei der I-S-T-Figurenauswahl FA (-.12), allerdings nicht so extrem wie bei den Würfelaufgaben WU (-.59).

Für Varimax: Die Varimax-rotierte Lösung enttäuscht, da der FRT nicht nur eine hohe Ladung auf Varimax-F_2 hat (.65), den die konventionelle Forschung als Faktor der ‚fluiden' Intelligenz interpretiert. Wider konventionelles Erwarten lädt der FRT mit .43 auch bei Varimax-F_1, dem Faktor der ‚kristallinen' Intelligenz, der beim FRT keine oder eine minimale Ladung haben sollte. Wider konventionelles Erwarten, aber mit Konsequenz, lädt auch die Figurenauswahl FA (.45) bei Varimax-F_1 hoch. Die Autoren selbst stellten die Ergebnisse einer Faktorenanalyse, bei der die I-S-T-Variablen zusammen mit FRT faktorisiert werden, nicht dar. Sie kommen aber mit einer alternativen Auswertung, bei der sie multiple Regressionen einsetzten (als UVs dienten ihnen die I-S-T Variablen, als AV der FRT-Gesamtwert) zu einem zwar komplizierteren, aber ebenso wenig überzeugenden Ergebnis.

Tabelle 3.07

Varimin- und Varimaxrotierte Faktorladungen des FRT (Nr. 9)
sowie Ladungen der I-S-T- 70-Subtestvariablen 1 - 8.

		Varimin		Varimax		
	Deutung	g	L_p	c	f	
	Subtests	F_1	F_2	F_1	F_2	h^2
1	*SE*	.63	.51	.80	.09	.65
2	*WA*	.63	.43	.75	.14	.59
3	*AN*	.77	.37	.81	.28	.73
4	*GE*	.75	.16	.65	.41	.58
5	*RA*	.76	.25	.72	.37	.66
6	*ZR*	.81	-.01	.56	.58	.65
7	*FA*	.76	-.12	.45	.62	.59
8	*WU*	.66	-.58	.05	.88	.77
9	*FRT*	**.76**	**-.15**	**.43**	**.65**	**.60**
	%Varianz	53	12	39	26	65

Anmerkung:
g = allgemeine Intelligenz L_p =Lernkapital
f = fluide c =kristalline Intelligenz

Zusammenfassung

Die Validitäten der Varimin- und Varimaxfaktoren, operationalisiert mithilfe von Tests, die Konvergenz-Divergenz-Urteile ermöglichen, fiel zugunsten von Varimin aus. Die Leistung bei zwei kulturunabhängigen Tests (CFT und FRT) wird von Varimin-F_1, dem Faktor der Grundintelligenz, weitaus besser differentiell vorhergesagt als von Varimax-F_2, dem üblicherweise die Bedeutung einer grundlegenden ‚fluiden' Intelligenz beigelegt wird.

Die Leistung bei zwei Außenkriterien (Rechtschreiben und Grundrechnen), bei denen sich Lernvorteile auswirken sollten, werden von Varimin-F_2, dem Faktor des Lernkapitals, besser differentiell vorhergesagt als von Varimax-F_1, dem die Bedeutung einer durch schulisches Lernen sich ausdrückenden ‚kristallinen' Intelligenz beigelegt wird.

Diskussion zu Kapitel 3

Die Untersuchung hatte zum Ziel zu prüfen, ob Varimin, das Verfahren der CSM-Rotation, auch auf dem Gebiet der Intelligenz die konventionelle SSM-Rotation an Effizienz übertrifft. Den Ergebnissen zufolge darf man diese Frage bejahen. Die aus dem I-S-T extrahierten Faktoren erwiesen sich nach einer Varimin-Rotation zur Aufdeckung latenter Varianzquellen als geeigneter als nach einer Varimax-Rotation.

Eine neue Forschungsstrategie hat, wenn sie gegenüber einer Konvention bestehen will, möglichst auch die Schwächen bisheriger Forschung aufzuzeigen. Ihre Hauptschwäche ist nach meiner Ansicht, dass Thurstones SSM-Modell, das bislang zugrunde gelegt wurde, die zu analysierenden Gegebenheiten fehlerhaft modelliert. Das Modell verschleiert, was Adolf Otto Jäger als „Kernannahme" an den Anfang aller Intelligenzforschung stellte: *„An jeder Intelligenzleistung sind (neben anderen Bedingungen) alle intellektuellen Fähigkeiten beteiligt, allerdings mit deutlich unterschiedlichen Gewichten. Die Varianz jeder Leistung lässt sich in entsprechende Komponenten zerlegen"* (Jäger ,1997, p.4). [87]

[87] Jäger hat sich bei seinen Datenanalysen vom SS-Prinzip zwar nicht getrennt, doch hat er methodische Alternativen anwenden müssen, um die in seinem BIS-Modell verwirklichten Gleichzeitigkeiten latenter funktionaler Beiträge zur Intelligenztestleistung empirisch zu belegen.

Durch Variminanalysen der Intelligenztest-Daten wurde deutlich, dass sich das schulische Lernen bei mangelnder Intelligenzausstattung bis zu einem gewissen Grade im Leistungsergebnis kompensatorisch auswirken kann. Diese Erkenntnis ist für die pädagogische Psychologie (Weinert, 1996) und auch für den *common sense* kein Novum. Neu ist allein die methodische Befreiung des Begriffs des Lerneinflusses vom Begriff der Intelligenz. Die beiden Begriffe werden in einer kristallinen Intelligenz, einem Artefakt der SSM, ohne Sachbegründung miteinander verschweißt. Durch CSM werden sie voneinander getrennt und sind dann operational und sachgerecht aufeinander beziehbar.

Es ließ sich zeigen, dass die SS-Rotation der Faktoren, die aus einem deplatzierten Sparsamkeitsmotiv heraus den Testvariablen nur einen Faktor (eine Varianzquelle) zuzuteilen sucht, den Hauptfaktor ‚*g*' zerstört. Dessen Beitrag zur Varianzaufklärung wird durch SSM auf alle nachfolgend extrahierten Faktoren aufgeteilt, die dadurch einen sachlich ungerechtfertigten hohen Zuschlag an Kommunalität erhalten. Dies war mit Tabelle 3.02 nachzuvollziehen. Ein Faktor F_3 wurde, wenn er Varimax-rotiert wurde, quantitativ aufgebläht – in den 18 Analysen auf unterschiedlichste Weise. Das gemittelte Endprodukt von Varimax-F_3 lieferte zwar bei den Subtests beachtliche Ladungshöhen, aber kein testdifferenzierendes Ladungsprofil, was eine inhaltliche Deutung von Varimax-F_3 ausschließt.

Sollte sich die zukünftige faktorenanalytische Forschung am Prinzip der Komplexstruktur orientieren, wird sie voraussichtlich mit viel weniger Konstrukten auskommen. Man wird Haupt- und Nebenfaktoren unterscheiden können, Initial-F_1 wird trotz Varimin-Rotation in der Regel weitgehend erhalten bleiben. Die auf Initial-F_1 folgenden Faktoren sind meist weniger varianzstark, ihre Kommunalitäten zeigen nach einer Varimintransformation relativ wenig Gewicht.

Die durch die weitere Forschung zu erwartenden Varimin-basierten faktoriellen Konstrukte werden vermutlich auch einfacher sein. Dies war bereits bei der Analye der I-S-T-Leistungen für Varimin-F_2 erkennbar. Man braucht zum Verständnis von F_2 keine neue Fähigkeit (‚kristalline Intelligenz') zu erfinden. F_2 lässt sich als Auswirkung eines biographisch erworbenen Lerngewinns verstehen, von dem die getesteten Probanden je nach Lernaufwand unterschiedlich profitieren, bei einigen Tests mehr, bei anderen weniger.

Als praktische Konsequenz der vorliegenden Ergebnisse könnte man den Autoren des IST für eine Neuauflage empfehlen, die Grundintelligenz (allgemeine Intelligenz *g*), die sie bislang in ihrem Test-Manual kaum angesprochen haben, über die der Test aber in erster Linie Auskunft geben kann (s. Varimin-F_1 bzw. I-S-T Gesamtwert), gebührend in den Vordergrund zu stellen. Information über das schulisch-kulturelle Lernkapital, die ebenfalls in den Testdaten steckt und bislang in verzerrter Form als eine besondere Sorte von Intelligenz behandelt wurde, könnte

die Validität des I-S-T ergänzen, allerdings anders als bisher. Man könnte die Rohwerte mithilfe von Varimin-F_2-Faktorgewichten transformieren und entsprechende Umrechnungstabellen zur Verfügung stellen. Einfacher wäre es, die Leistungen in den Tests, bei denen sich das Lernkapital stärker auswirkt, zur Gesamttestleistung ins Verhältnis zu setzen (durch L_D oder L_p), wodurch der relative Beitrag der schulisch-kulturellen Erfahrung zur I-S-T Gesamtleistung für die einzelnen Testteilnehmer abschätzbar würde.

Kapitel 4

Komplexität sportlicher Leistungen.

Generelle sportliche Kondition, Ort des Energieeinsatzes,
Dosierung des Energieeinsatzes.

Vorspann

Zu den historischen Anfängen der faktorenanalytischen Methode stellen sich schwer beantwortbare Fragen: Warum hat man das Funktionieren der neuen Methode fast immer nur mit Variablen der Psychologie zu überprüfen versucht (mentale Leistungen, Selbst- und Fremdbeurteilungen)? Was immer beim Bearbeiten von Denkaufgaben und beim Verbalisieren von Verhaltensdispositionen (von ‚traits') in uns abläuft, ist unserer Wahrnehmung nicht unmittelbar gegeben. Man ist auf Erlebnisbeobachtungen angewiesen, die sprachlich-semantisch verarbeitet werden und zu unpräzisen und unsicheren Urteilen führen. Hätte man zur Überprüfung der Faktorenanalyse als Methode zunächst Datensätze mit Variablen herangezogen, über deren Bedeutung man sich leichter einigen kann, etwa Variablen der körperlichen Leistung, dann hätte man zur Frage, wie die Güte der neuen Methode einzuschäzen ist, wahrscheinlich eher einen Konsens erreicht.

Solche Überlegungen brachten mich auf den Gedanken, die Domäne der sportlichen Leistungen einer faktoriellen Untersuchung zu unterziehen. Erfreulicherweise gibt es die Zehnkampfdisziplin, die den Sportlern verschiedenartige körperliche Leistungen abverlangt. Zudem sind wertvolle Datensätze mit olympischen Zehnkampfleistungen öffentlich zugänglich. Merkwürdig ist, dass diese in den Lehrbüchern zur Faktorenanalyse nicht auch als Demonstrationsbeispiele verwendet werden. Ich vermute, der Grund ist das unbefriedigende, weil kaum interpretierbare Ergebnis einer Faktorenanalyse dieser Daten, das mit Rotation zur Einfachstruktur entsteht (s. unten). Die Interpretation von komplexstrukturell rotierten Faktoren dieser Daten besticht dagegen durch Einfachheit und Transparenz. Vorweggenommen sei das Hauptergebnis: Die Leistungen von Zehnkämpfern variieren hinsichtlich der Gesamtleistung (Generalfaktor ‚g', F_1), sie variieren auch

zwischen einer höheren Beanspruchung der Motorik des oberen gegenüber dem unteren Extremitätenbereich (F_2). Unabhängig davon variiert die zeitliche Verteilung des Energie-Einsatzes, wofür Sportfachleute seit langem die Polarität Schnellkraft vs. Durchhaltekraft eingeführt haben (F_3). An diese Deutung der Leistungsvarianzen im Zehnkampf, die sich durch die komplexstrukturelle Faktorenanalyse ergibt, lassen sich nahtlos sportphysiologische Erkenntnisse anknüpfen, das wird im Bericht weiter ausgeführt.

Ausgangslage und Zielsetzung

Vom Unbehagen der exploratorisch-faktorenanalytischen Forschung (EFA), das sich hierzulande spätestens seit dem ‚Münchener Symposium' (Kallina, 1968) artikuliert hat, blieben auch sportpsychologische Untersuchungen nicht verschont (Bös, 1987, Teipel, 1988, p. 341 ff., Büsch et al. 2001). [88] Da die Enttäuschungen nicht nachlassen, greift man zunehmend zu neueren Entwicklungen der multivariaten Methodik. Mit konfirmatorischen Faktorenanalysen und Strukturgleichungsverfahren werden vermutete latente Parameter der Variablen einer sukzessiven Prüfung auf Übereinstimmung mit der Datenlage unterzogen.

Ob sich mit dieser Forschung die Ergebnisbilanz verbessern wird, bezweifeln u.a. Cliff (1983), Church & Burke (1994), Ferrando & Lorenzo-Seva (2000). Konfirmatorisch modell-testendes Forschen kann ein exploratorisches Aufspüren latenter Bedingungen nicht ersetzen, das Hypothesentesten bewegt sich im Bereich des Erwartbaren. Unerwartete Befunde lassen sich dann kaum gewinnen. Überdies hat man es versäumt, der Frage auf den Grund zu gehen, warum die exploratorischen Verfahren bisheriger Konvention nicht das geleistet haben, was sie leisten sollten.

---

[88] Zur Beurteilung der *faktoriellen Validität* sportmotorischer Tests, die Bös (1987) in einem Handbuch darstellt, meint der Verfasser (p. 141): *„Eine detaillierte Analyse der zitierten [faktorenanalytischen] Befunde lässt zum Teil eine ‚erschreckende Unverbindlichkeit' (Orlik, 1967, p. 87) der faktorenanalytischen Dimensionsanalysen erkennen."* Die Ergebnisse führten oft zu „Widersprüchen", lassen „keine eindeutigen Interpretationen zu", „die Aussagekraft ...(werde) überschätzt". Probleme lägen schon „in der generellen Anwendbarkeit der Faktorenanalyse". Seit längerem gebe es „Zweifel an der Tragfähigkeit faktorenanalytischer Befunde". Faktorenstrukturen können „nur selten repliziert werden…" (p. 461).

Dieser Frage wurden inzwischen mehrere Untersuchungen gewidmet (Kapitel 2 und 3 und Ertel a, b). Nach meinem Dafürhalten sind durch *Simple Structure Modeling* (SSM), nach welchem sich die EFA mit all ihren Verzweigungen seit Thurstone (1935, 1947) ausgerichtet hat, die Weichen falsch gestellt worden. SSM ist eine rein mathematische Richtschnur zur Transformation von Faktoren. Sie führt dazu, dass den manifesten Variablen ein Minimum an aufgefundenen Faktoren, im Idealfall nur ein Faktor pro Variable zugeteilt wird. Tatsächlich beobachtet man aber in der Psychologie wie auf allen Ebenen der Natur, dass einzelnen Observablen nicht einzelne oder möglichst wenige, sondern multiple latente Bedingungen zugrunde liegen, die gleichzeitig zusammenwirken. Was Thurstone aus mathematischer Sicht *Einfachstruktur* nennen konnte, hat durch stillschweigende Übertragung auf die Ebene empirischer Gegebenheiten, die durch den SSM-Algorithmus abgebildet werden sollen, zum Gegenteil geführt: zu zusammenhanglosen Aggregaten.

Die seit langem als Standard verwendete SSM Rotation *Varimax* wurde deshalb gegen das hier neu entwickelte Verfahren *Varimin* ausgetauscht. Durch Varimin werden die Faktor-Koordinaten so gedreht, dass die manifesten Variablen, die der Analyse unterworfen werden, an möglichst allen faktoriell ermittelten latenten Variablen partizipieren, soweit dies die Datenlage ermöglicht. Varimin orientiert sich am *Complex Structure-Model* (CSM), der Antithese des SSM, mit der Erwartung, dass das nach der Richtschnur der CS modellierte Ergebnis die Realität angemessener und einfacher abbildet als das nach SS modellierte Ergebnis. Varimax maximiert die Varianz der Ladungsquadrate der Faktoren über die Variablen, durch Varimin werden die Ladungsquadrate minimiert. Dass von CSM einfachere und verständlichere (nicht kompliziertere) Ergebnisse erwartet werden als vom SSM, ist nur scheinbar paradox, diese Erwartung erweist sich bei näherer Prüfung als wohlbegründet (s. Kapitel 2).

Die SSM-orientierte Transformation initialer Faktoren mit Varimax als Standardverfahren, die sich zum ‚Ritual' der EFA-anwendenden Forscher entwickelt hat (Gigerenzer & Strube, 1978), wird sich indessen mit theoretischen Argumenten allein kaum erschüttern lassen. Hinzu kommen muss der Nachweis, dass die Ergebnisse der CSM-orientierten Transformation stabiler und valider sind als die Ergebnisse der SSM-orientierten Transformation. [89]

Zur Prüfung dieser Frage eignen sich insbesondere Daten, bei denen die Bedingungsparameter a priori bekannt sind (s. Ertel, 2009b). Auch sind Daten aus Do-

[89] Erste Bewährungsproben hat das Varimin-Verfahren auf dem Gebiet der Persönlichkeitstraits (Ertel, a) und der Intelligenz (Ertel, b und Kapitel 3) zu bestehen gehabt.

mänen des Verhaltens geeignet, deren latente Bedingungen begrifflich leicht zu fassen sind. Bei der Deutung von Faktoren körperlicher Leistung dürften Meinungsverschiedenheiten weniger häufig auftreten als bei der Deutung von Faktoren von Daten intellektueller Leistungen, bei denen die Testpersonen perzeptiv nicht-zugängliche Denkoperationen mit Wörtern, Sätzen und Zahlen aktivieren.
Der Zehnkampf (engl. decathlon) ist eine Sportkampfform, bei der in zehn verschiedenen sportlichen Disziplinen eine Vielzahl körperlicher Leistungen verlangt werden. Da die Leistungen in diesen Disziplinen untereinander alle, z. T. hoch positiv, korrelieren (s. Interkorrelationen in Tabelle 4.01a und 4.01b), ist mit <10 zugrunde liegenden Leistungsfaktoren zu rechnen. Von einer mit CSM durchgeführten Faktorenanalyse wird zunächst ein sportsspezifischer ‚g'-Faktor erwartet analog einem ‚g'-Faktor der Intelligenz, und dies mit analoger Begründung (Jensen, 1998). Ein ‚g'-Faktor würde der anlagemäßig gegebenen sowie durch Training verbesserten Leistungsdisposition entsprechen, die bei Sportlern verschiedene Ausprägungsgrade erreicht (‚Allgemeindisposition'). Doch werden ergänzende Faktoren hinzukommen. Nach Linden (1977), der bislang offenbar die einzige faktorenanalytische Untersuchung zum Zehnkampf vorgelegt hat (beschrieben von Basilevsky, 1994), ist mit etwa vier Faktoren zu rechnen. Linden hat sie nach Varimax transformiert und gedeutet als F_1: *Kurzstrecken-Schnell-Lauf*, F_2: *Explosive Armkraft*, F_3: *Ausdauer beim Lauf*, F_4: *Explosive Beinkraft*.

Kritischer Einschub: Ein ‚g'-Faktor der allgemeinen sportlichen Disposition (Anlage plus Trainingsgewinn) taucht in Lindens Varimax-Lösung nicht auf. Die von ihm verwendeten Bezeichnungen lassen mit ihrer Spezifität vermuten, dass der Forscher die Ebene latenter Leistungskomponenten nicht erreicht hat: Warum sollen sich z. B. Armkraft und Beinkraft nur ‚explosiv' ausdrücken können? Warum sollen sich Fähigkeiten für schnelles Laufen nur bei Kurzstrecken manifestieren? Der Stabhochsprung wird von den Beschreibungskategorien Lindens überhaupt nicht berührt, er passt nicht in sein Faktorensystem.

Drei weitere Decathlon-Datensätze wurden gefunden und zur Absicherung der Ergebnisse mit herangezogen: Eine Interkorrelationstabelle von Kunz (1980), Wettkämpfer-Originaldaten von Zarnowski (1989) und ebenfalls Originaldaten aus einer Internetquelle (2004). Die Interkorrelationen der Leistungen zwischen den zehn Sportarten, gewonnen aus den vier Quellen, werden in Tabelle 4.01a und 4.01b wiedergegeben.

Tabelle 4.01a

Interkorrelationen von Zehnkampf-Leistungen aus den Quellen Linden und Kunz

	1 100 m Lauf	2 Weit- sprun g	3 Ku- gel- wurf	4 Hoch- sprun g	5 400 m Lauf	6 110m Hür- den	7 Dis- kus- Wurf	8 Speer - Wurf	9 Stabh. - sprung	10 1500 m Lauf
1	-	.59	.35	.34	.63	.40	.28	.20	.11	-.07
2	.50	-	.42	.51	.49	.52	.31	.36	.21	.09
3	.15		-	.38	.19	.36	.73	.24	.44	-.08
4	.17	.38	.11	-	.29	.46	.27	.39	.17	.18
5	.68	.20	-.01	.27	-	.34	.17	.23	.13	.39
6	.55	.33	.28	.36	.55	-	.32	.33	.18	.00
7	.15	.50	.54	.16	.11	.31	-	.24	.34	-.02
8	.16	.39	.16	.28	.39	.35	.28	-	.24	.17
9	-	.37	.27	.01	-	.02	.33	.06	-	.00
1	.19	.23	-.13	.10	.13	.34	.07	.33	.19	-
0	.25	.19			.62					

Anmerkung:
Obere Dreiecksmatrix: Quelle Linden (1977). Leistungen auf den Olympiaden 1948-1976.
Untere Dreiecksmatrix: Quelle Kunz (1980) . Leistungen Schweizer Spitzensportler. Die Vorzeichen der Korrelationen zwischen den Laufdisziplinen, bei denen sich höhere Leistung durch kürzere Zeitmaße ausdrückt, und den übrigen Disziplinen, sind im Original negativ.

Tabelle 4.01b

Interkorrelationen von Zehnkampf-Leistungen
aus den Quellen Website und Zarnowski.

	1 100m Lauf	2 Weit- sprung	3 Kugel- wurf	4 Hoch- sprung	5 400m Lauf	6 110m Hürden	7 Diskus- Wurf	8 Speer- Wurf	9 Stabh.- sprung	10 1500m Lauf
1	-	.58	.31	.15	.64	.52	.25	.14	.12	.11
2	.66	-	.38	.40	.39	.55	.34	.49	.18	.15
3	.51	.56	-	.20	.18	.46	.79	.49	.60	-.03
4	.45	.64	.50	-	.17	.34	.16	.18	.26	.09
5	.66	.60	.39	.54	-	.43	.15	.24	.11	.52
6	.62	.67	.55	.63	.58	-	.48	.43	.13	-.07
7	.43	.48	.80	.44	.37	.49	-	.42	.47	-.05
8	.48	.61	.56	.70	.59	.63	.52	-	.31	.15
9	.34	.46	.59	.40	.44	.41	.51	.52	-	-.04
10	.09	.21	.09	.35	.53	.21	.12	.37	.24	-

Anmerkung:
Obere Datenmatrix: Quelle Website: Österreichische Spitzenleistungen (2400)
Untere Dreiecksmatrix: Quelle Zarnowski (1989): Leistungen auf den Olympiaden 1948-1988. Zur Vorzeichenbehandlung siehe Anmerkung zu Abbildung 1a.

Datenbeschreibung

1. Linden (1977)[90]

Die Korrelationstabelle von Linden stützt sich auf olympische Zehnkampfleistungen von n = 160 Wettkämpfern, die auf acht Olympiaden (1948-1976) erbracht wurden. Zum Aufweis latenter Bedingungen verwendet Linden die Faktorenanalyse. In der vorliegenden Untersuchung wurden die Daten von Linden verwendet, die durch Basilevsky (1994) mitgeteilt wurden.

2. Kunz (1980)
Die Korrelationstabelle von Kunz (S. 161) stützt sich auf Zehnkampfleistungen von n = 27 Schweizer Zehnkämpfern der Spitzenklasse, die ihre Leistungen auf 8 Wettkämpfen erbrachten (die Datenbasis beinhaltete 90 Wettkampfresultate, da viele Zehnkämpfer sich an mehreren Wettkämpfen beteiligten). Kunz interpretiert die n(n-1)/2 Interkorrelationen, eine Faktorenanalyse verwendet er nicht. Die Rohdaten, die der Korrelationstabelle zugrunde liegen, teilt Kunz in einem Appendix mit (S. 212/13).

3. Zarnowski (1998)
Dieser Quelle ließen sich die Rohdaten von 233 Zehnkämpfern entnehmen, die ihre Leistungen auf 11 Olympiaden (1948-1988) erbrachten. Die Leistungen von 75 Wettkämpfern, die nicht alle zehn Disziplinen absolvierten, blieben unberücksichtigt. Die Daten der Quelle *Linden* sind in den Daten der Quelle *Zarnowski* enthalten.

4. http://www.zehnkampf (2000). Dieser Websites österreichischen Ursprungs wurden Rohdaten von N = 108 österreichischen *Alltime Decathlon* Kämpfern entnommen, darunter befinden sich N = 64 *Alltime Decathlon Juniors* (Alter < 20). (Über die URL http://werthner.casc.at/bin/results/alltime.php sind Daten von mehreren Tausend österreichischen Zehnkämpfern zugänglich, was erst nach Ab-

[90] Die Daten von Linden werden wegen ihrer Anschaulichkeit gelegentlich für Übungszwecke in Statistik-Kursen verwendet und sind über das Internet zugänglich, derzeit z. B über http://math.usask.ca/~miket/f-03.pdf

schluss der Untersuchung festgestellt wurde, siehe letzte Fußnote am Ende dieses Kapitels).

Datenverarbeitung: Faktorenanalysen

Der Verlauf der sechs ersten Eigenwerte bei den vier Analysen (PCA) ist der Tabelle 4.02 zu entnehmen. Nach dem Kaiser-Guttman-Kriterium wären drei Faktoren für eine Rotation zu berücksichtigen, auch wenn man mit diesem Kriterium allein bei den Zarnowski-Daten eine andere Entscheidung treffen könnte. Bei den Zarnowski-Daten würde indessen auch die Interpretierbarkeit der Faktoren für eine Dreierlösung sprechen.

Tabelle 4.02

Verlauf der Eigenwerte 1 bis 6 bei den vier Hauptkomponenten-Analysen

	1	**2**	**3**	**4**	**5**	**6**
Linden	3.79	1.52	1.11	0.91	0.72	0.60
Kunz	4.43	1.85	1.18	0.96	0.71	0.59
Zarnowski	5.44	1.24	0.92	0.63	0.50	0.33
Website	3.88	1.76	1.08	0.96	0.80	0.53
a. M.	**4.39**	**1.59**	**1.07**	**0.87**	**0.68**	**0.51**

Die vier Initial-Lösungen der PCA wurden nach Varimin und zum Vergleich auch nach Varimax transformiert. Zunächst sollte ermittelt werden, ob, wie vermutet wurde, die Varimin-Lösungen größere faktorielle Übereinstimmungen unter den vier Datensätzen aufweisen als die Varimax-Lösungen. Die Kongruenzkoeffizienten nach Tucker-Wrigley & Neuhaus (s. Harmann, 1968, p. 270) sind Tabelle 4.03 zu entnehmen.

Tabelle 4.03

Kongruenzen der rotierten Faktoren F_1, F_2, und F_3, differenziert nach vier Datenquellen. Die drei jeweils niedrigsten Kongruenzen für F_1, F_2, und F_3 sind für jeden Faktor fett gedruckt.

	F_1		F_2		F_3	
	Varimin	*Varimax*	*Varimin*	*Varimax*	*Varimin*	*Varimax*
Linden : Kunz	.98	**.93**	.94	**.66**	.94	**.79**
Linden : Zarnowski	1.00	.99	.97	.97	.98	.97
Linden : Website	.98	.99	.98	.98	.92	.88
Kunz : Zarnowski	.95	**.94**	.93	**.53**	.87	.89
Kunz : Website	.98	**.90**	.90	**.63**	**.79**	**.82**
Zarnowski : Website	.98	.98	.95	.97	.96	.89
a. M.	**.978**	**.955**	**.945**	**.790**	**.910**	**.873**

Bei den Varimax-Lösungen finden sich bei jedem Faktor (F_1, F_2 und F_3) geringe Kongruenzen zwischen den Datenquellen. Die Kongruenzen (a. M.) liegen bei den *Varimin*-Lösungen höher, diese erweisen sich somit gegenüber dem Wechsel der Datenquellen als stabiler.

Bei den vorliegenden hohen Kongruenzen der Varimin-rotierten Faktoren verschiedener Datensätze genügt es, zur Faktorendeutung die Ergebnisse nur eines Datensatzes zu verwenden. Die Quelle Zarnowski erscheint dafür am geeignetsten, weil sie das höchste N der Zehnkämpfer aufweist, die aus 49 verschiedenen Nationen stammen. Zudem enthält *Zarnowski* die Daten der weniger umfasssenden Quelle *Linden*. Die Ausgangsdaten von Zarnowski sind Originalleistungen, zehn pro Wettkämpfer pro Olympiade. Die sieben besten Olympia-Leistungen sind beispielhaft in Tabelle 4.04 wieder gegeben.

Tabelle 4.04

Die sieben besten Olympialeistungen im Zehnkampf nach offizieller Gesamtpunktzahl aus der Quelle *Zarnowski.*

Jahr	Ort	Rang	Nation	Name des Sportlers	100m	Weitsprg	Kugel	Hochsprg	400m	110m Hürd	Diskus	Stabhoch	Speer	1500m	Gesamtpunkte
1984	Los Ang	1	GB	Daley Thompson	10.44	801	1572	203	46.97	14.33	46.56	500	65.24	4:35.00	8847
1984	Los Ang	2	D	Jürgen Hingsen	10.91	780	1587	212	50.82	14.29	40.82	450	60.44	4:22.60	8695
1976	Montreal	1	USA	Bruce Jenner	10.94	722	1535	203	50.04	14.84	50.04	480	68.52	4:12.61	8634
1988	Seoul	1	D	Christian Schenk	11.25	743	1548	227	49.28	14.13	49.28	470	61.32	4:28.95	8488
1972	München	1	SU	Nikolay Avilov	11.00	768	1436	212	46.98	14.31	46.98	455	61.66	4:22.80	8466
1976	Montreal	2	D	Guido Kratschmer	10.66	739	1474	203	45.70	14.58	45.70	460	66.32	4:29.09	8416
1984	Los Ang	3	D	Siegfried Wentz	10.99	711	1587	209	46.60	14.35	46.60	450	67.68	4:33.96	8416

Manche Wettkämpfer nahmen an mehr als einer Olympiade teil. Die Faktorladungen der Varimin-Lösung (Zarnowski-Daten) sind zusammen mit den Ladungen der Varimax-Lösung und mit der Initial-Lösung in Tabelle 4.05 zu finden.

Tabelle 4.05

Varimin-, Varimax- und unrotierte Faktorgewichte der zehn Sportdisziplinen des Zehnkampfes (Datenquelle Zarnowski: N = 233 Olympioniken).

	Aufgabenart	*Varimin-Lösung*				*Varimax-Lösung*				*Initial-Lösung*			
		F_1	F_2	F_3	h^2	F_1	F_2	F_3	h^2	F_1	F_2	F_3	h^2
1	**100 m Lauf**	**.49**	**.61**	**.42**	.79	**.22**	**- .03**	**.86**	.79	**.73**	**-.07**	**.50**	.79
2	Weitsprung	.66	.47	.30	.75	.36	.15	.77	.75	.82	-.01	.27	.75
3	Kugelstoß	.69	.51	-.37	.86	.86	-.01	.35	.86	.77	-.46	-.22	.86
4	Hochsprung	.74	.18	.23	.64	.37	.40	.59	.64	.78	.17	.03	.64
5	**400 m Lauf**	**.75**	**.06**	**.46**	.78	**.17**	**.54**	**.67**	.78	**.77**	**.41**	**.13**	.78
6	**110 m Hürden**	**.65**	**.47**	**.30**	.75	**.36**	**.15**	**.76**	.75	**.81**	**-.01**	**.27**	.75
7	Diskuswurf	.66	.43	-.42	.79	.85	.02	.26	.79	.71	-.45	-.30	.79
8	Stabhochsprung	.82	.16	.10	.67	.50	.44	.52	.67	.82	.12	-.11	.67
9	Speerwurf	.73	.11	-.35	.71	.74	.31	.15	.71	.66	-.15	-.45	.71
10	**1500 m Lauf**	**.65**	**-.65**	**.23**	.89	**.02**	**.94**	**.06**	.89	**.39**	**.76**	**-.40**	.89
	% Varianz	**47.2**	**17.2**	**11.2**	**76.0**	**27.2**	**16.7**	**32.0**	**76.0**	**54.4**	**12.4**	**9.2**	**76.0**

Anmerkung: Die Daten der vier Laufdisziplinen wurden zur Verbesserung der Übersichtlichkeit fett wiedergegeben

1. Deutung der Variminfaktoren.

F_1 ist der erwartete Generalfaktor *g* der sportlichen Leistungen (Allgemeinkondition). Bei der sportlichen Kondition als Allgemeinbedingung für sportliche Leistungen kann mangels differenzierender Testleistungen nicht – wie bei den intellektuellen Testleistungen - zwischen Anlage- und Lern- oder Traininganteilen unterschieden werden. Die Faktorscores von $\mathbf{F_1}$ bei den 233 Sportlern korrelieren hoch mit der Decathlon-Gesamtleistung (r = .95). Die individuellen Gesamtleistungen, die nach der Vorschrift des Sportverbandes offiziell berechnet werden, sind gewichtete Summen der Leistungen in den zehn Disziplinen [91]. Die Werte der Gesamtleistung stammen aus der Quelle Zarnowski, sie wurden nicht auch als Variable mit faktorisiert. Die F_1 -Faktorscores sind also aufgrund ihrer hohen Korrelation mit dem Gesamtwert als unabhängig validiert zu betrachten.

Leistungen bei Sportarten, die ein relativ hohes Gewicht bei F_1 aufweisen (z. B. Stabhochsprung), haben weniger F_2 und F_3-Bezüge als Sportarten, die ein weniger

[91] Das Punkte-System für den Zehnkampf wurde 1985 von einem Komitee der IAAF festgesetzt. Eine mathematische oder statistische Begründung der Bewertungen im einzelnen wurde nicht mit veröffentlicht.

hohes Gewicht bei F_1 aufweisen (z. B. 100 m-Lauf). Kunz erwähnt auffällige korrelative Beziehungen zwischen Stabhochsprung und den Lauf-, Sprung- und Wurfdisziplinen und meint, dass dies auf eine „Vielseitigkeit dieser Disziplin hinweist“ (p. 166). Vielseitigkeit impliziert Abhängigkeit von einer größeren Zahl von Bedingungen, die sich bei ausgewogener Ausprägung in vermehrter *g*-Faktor-Ladung geltend machen kann.

Varimin-F_2 ist ein bipolarer Faktor, der beim 1500 m-Lauf eine sehr hohe Ladung aufweist, sie hat als einzige ein negatives Vorzeichen (-.65). Zur Deutung von Varimin-Faktoren ist die Anwendung des *minimal pair-Vergleichs* oft unverzichtbar (s. Kapitel 2). Im vorliegenden Fall lässt sich der Vergleich beispielhaft auf den 1500 m-Lauf im Verhältnis zum 100 m-Lauf anwenden. Die F1, F2, F3 Ladungsprofile:

	F_1	F_2	F_3
1500 m-Lauf	.65	**-65**	.23
100 m-Lauf	.49	**.61**	.42

zeigen einen starken Kontrast bei F_2, während sich die Ausprägungen von F_1 und F_3 nur geringfügig unterscheiden. Ein Ladungskontrast bei jeweils nur einem Faktor ist für Minimalpaar-Vergleiche notwendig. Offensichtlich wird also beim Vergleichspaar 1500 m- vs. 100 m-Lauf in der kontrastierenden F_2 –Ladung ein polar aufzufassender Bedingungskomplex manifest. Es liegt nahe, den Begriff ‚Ausdauer‘, den die Sportsprache verwendet, hier zu übernehmen (eine ältere Alternativbezeichnung ist ‚Stehvermögen‘, Kunz, 1980). ‚Ausdauer‘ hat ein polares Gegenstück in dem, was in der Sportsprache ‚Schnellkraft‘ heißt. Beim 1500 m-Lauf kommt es auf eine ausdauernde Kraftentfaltung an, beim 100 m-Lauf auf eine schnellere Kraftentfaltung. Von den Laufdisziplinen zeigt plausiblerweise auch der 110 m-Hürden-Lauf Schnellkraft-Einsatz (F_2: .47), während der 400 m-Lauf auf der bipolaren F_2-Skala Skala mit F_2 = .06, ebenfalls plausiblerweise, zwischen dem 100 m- und dem 1500 m-Lauf zu liegen kommt. Bei den Wurf-Disziplinen sollten nach dem Faktoren-Ergebnis das Kugelstoßen (F_2 = .51) und der Diskuswurf (.43) mehr Schnellkraft erfordern als der Speerwurf (.11). Der Speerwurf hat nach Kunz „vielseitigere Voraussetzungen“, bei dieser Disziplin ist „wahrscheinlich die außerordentlich anspruchsvolle Technik ausschlaggebend“ (Kunz, p. 167), wodurch sich der Speerwurf von den typischen Schnellkraft-Disziplinen unterscheidet – bei den Wurf-Sportarten sind Schnellkraft-Prototypen der Kugelstoß und der Diskuswurf.

Insofern die Deutung von F_2 zutrifft, kann man damit einverstanden sein, dass unter den Sprung-Disziplinen der Weitsprung eine erhöhte Schnellkraft voraussetzt (F_2 = .47), während der Hochsprung (.18) und der Stabhochsprung (.16) weniger Schnellkraft zu erfordern scheinen, was ebenfalls nicht unplausibel ist.

Mit F_2 (Varimin) werden also offenbar Unterschiede im Tempo der Verausgabung an Körperkraft manifest. [92]

F_3 ist ebenfalls ein bipolarer Faktor. Für diesen sind einerseits die Sportarten charakteristisch, die insbesondere Kraftentfaltung in den oberen Extremitäten erfordern, das sind die Wurf-Disziplinen Diskuswurf (F_3= -.42, Speerwurf (F_3 = -.35) und das Kugelstoßen (F_3 = -.37). Hinzu kommen bei diesem Faktor mit gegensätzlichem Vorzeichen die Sportarten, die insbesondere Kraft der unteren Extremitäten erfordern, das sind die Laufdisziplinen mit den Strecken 400 m (F_3 = .46), 100 m (F_3 = .42), und 110 m Hürden (F_3 =.30). Der Langlauf über 1500 m erfordert offensichtlich nicht primär Beinkraft (F_3 = .23) als vielmehr Ausdauer des Energie-Einsatzes (F_2).

Die hier vorgelegte Deutung der Faktoren ist gekennzeichnet durch Einfachheit. Haupt-Varianzquellen der Zehnkampfleistungen neben dem *g* -Faktor ergeben sich durch örtliche und zeitliche Bedingungen, d. h. der ‚Ort' der Kraftentfaltung (oberes Bewegungssystem vs. unteres Bewegungssystem) und die zeitliche Erstreckung der eingesetzten Kräfte (zeitlich konzentriert vs. gestreckt) sind entscheidend.

Das Kategorienschema, basierend auf empirischen Korrelationen, ist für die sportphysiologische Forschung und ihre Anwendung geeignet. So kann man z. B. zur Generalisierung des Trainingserfolgs für verschiedene Zehnkampf-Sportarten schlussfolgernde Überlegungen anstellen (was gewinnt Sportart B durch Training der Sportart A?). Im Augenblick genügt es zu konstatieren, dass ein leicht verständliches Endergebnis erzielt wurde, und dass dies deshalb möglich war, weil der Manifestation eines komplexen Faktorenmusters der Weg geebnet wurde. Der 100 m-Lauf z. B. wird nicht wie bei Linden einer einzigen Kategorie zugeordnet (*‚Kurzstrecken-Schnell-Lauf'*), vielmehr werden Ausprägungswerte der drei ermittelten Faktoren berücksichtigt: Hohe Leistungen beim 100 m-Lauf erfordern ein möglichst hohes allgemeines körperliches Leistungsvermögen (physische Konstitution, Anlage plus Trainingsgewinn machen die Kondition aus). Sie erfordern zudem erhöhte Beinkraft und den geballten Einsatz körperlicher Energien (Schnellkraft). Auch die anderen Sportarten lassen sich durch variable Ausprägungswerte der gleichen Faktoren charakterisieren, erst durch das Heranziehen gleicher latenter Merkmalkategorien werden sie miteinander vergleichbar. [93]

---

[92] Für Schnellkraft-Ausdauer (F_2) bei sportlichen Leistungen gibt es bei mentalen Leistungen eine analoge Polarität (speed - power), die in der Intelligenzforschung diskutiert und empirisch untersucht wird (Jensen,1993, pp.492-509).

[93] Dabei ist im Auge zu behalten, dass mit Auswertung von Leistungsunterschieden, die zwischen Sportlern auftreten, von den leistungsbedingenden Pa-

2. Deutungsversuch für die Varimaxfaktoren.

Die drei *Varimax*faktoren haben fast nur positive Ladungen, sind also unipolar, was analog bei Intelligenztest-Analysen die Regel ist. Ein Generalfaktor tritt wie bei faktoriellen Intelligenzanalysen nicht in Erscheinung. Die Faktorscores der 233 Olympioniken korrelieren mit den offiziellen Gesamtpunktwerten ihrer Leistungen, also mit dem Außenkriterium des zu fordernden Generalfaktors, wie folgt: r = .60 (für F_1), r = -.45 (für F_2) und r = -.64 (für F_3). Die Zusammenhänge sind hochsignifikant, aber inkonsistent, ungefähr so, wie man es von der durch Varimax hervorgerufenen Verteilung der Gesamtleistungsvarianz auf drei Faktoren zu erwarten hat.[94] Schon nach modelleigenen Maßstäben ist das Varimax-Ergebnis fehlerbelastet, denn eine Einfachstruktur findet sich nur bei drei der zehn Sportvariablen. Sieben Disziplinen sind aus Sicht der SS als unwillkommene Mischformen hinzunehmen. [95]

Was geben die Varimax-Faktoren inhaltlich her? F_1 mit höchsten Ladungen beim Kugelstoßen (.86), Diskuswurf (.85), und Speerwurf (.74) sieht zunächst aus wie ein Faktor der Wurfsportarten. Doch hat auch der Stabhochsprung eine ziemlich hohe Ladung bei F_1 (.51) und es fällt schwer, beim Stabhochsprung eine Ähnlichkeit zu den Wurfsportarten zu sehen. Auch beim Weitsprung und Hochsprung liegen die F_1-Ladungen noch über .30, was mit einer Deutung von F_1 als wurfsportart-zugehörig oder wurfsportart-ähnlich kaum vereinbar ist.

---

rametern nur diejenigen identifizierbar werden, die diesen Unterschieden zugrunde liegen. Die nicht oder nur geringfügig zwischen Individuen variierenden latenten Parameter bleiben latent und können nur über andere methodische Zugänge Gegenstand der Forschung werden.

94 Dass die Vorzeichen von zwei Koeffizienten negativ sind, darf man nicht als erwartungswidrig interpretieren, denn die Vorzeichen der Faktorenladungen sind oft beliebig. Die verfälschende Wirkung, die eine SSM-Rotation von Sporttest-Daten hervorruft, beklagen auch Highmore & Taylor (1954): *„…the,'basic' factor, representing general athletic ability (in which we are primarily interested), necessarily disappears, and the group factors[of simple structure rotation] show little relation to the classification indicated by the [initial] bipolar matrix"* (p. 4).

95 Schon der Klassiker Harry Harman musste zugeben: *"An orthogonal uni-factor solution is practically impossible with empirical data, and not very likely even when the factors are permitted to be oblique. Nonetheless, … it is towards that end that the simple structure principles are proposed for the multiple factor solution"* (Harman, 1968, p. 99).

Varimax-F_2 mit höchster Ladung beim 1500 m-Lauf (.94) scheint zunächst für ‚Ausdauer' zu sprechen. Auch passt der 400 m-Lauf mit F_2 = .54 noch zu dieser Deutung. Doch relativ hohe Ladungen von F_2 finden sich auch beim Stabhochsprung (.44) und Hochsprung (.40), was dieser Deutung widerspricht. Eine Gemeinsamkeit zwischen dem 1500 m-Lauf und den beiden Hochsprung-Disziplinen wird man schwerlich behaupten können.

Varimax-F_3 mit höchsten Ladungen beim 100 m-Lauf (.86), auch beim Weitsprung (.77), sowie bei 110 m-Hürden (.76), beim 400 m-Lauf (.67) und auch beim Hochsprung (.59) mischt Lauf- und Sprung-Disziplinen durcheinander, eine sinnvolle Deutungsmöglichkeit ist bei F_3 kaum gegeben.[96]

3. Deutungsversuch für die Initialfaktoren.

Man kann den Verdacht haben, Varimin-Faktoren seien mit den Initialfaktoren identisch oder Initiallösungen könnten, wo sie von den Variminlösungen abweichen, hinsichtlich Deutbarkeit mit diesen möglicherweise konkurrieren. Der vorliegende Fall widerspricht diesem Verdacht. Zwar zeigt sich eine sehr hohe Kongruenz (Tucker-Wrigley-Neuhaus) zwischen Varimin-F_1 und Initial-F_1 .98 (s. Tabelle 4.06b). Doch für F_2 beträgt sie nur noch |.62|. und für F_3 .69. Von einer

---

[96] Nach Abschluss des Manuskripts wurde eine umfassendere österreichische Datenquelle mit 4,586 vollständigen Zehnkampfleistungen von 2,674 Sportlern entdeckt (http://werthner.casc.at/bin/results/), dort in einer ‚Alltime-Liste Wien. 1993-2004' zu finden). Die Olympia-Leistungen, die in der vorliegenden Arbeit zur Faktorendeutung verwendet wurden, zeigen gegenüber den Leistungen der Quelle 'Wien 1993-2004' hohe faktorielle Kongruenzwerte: Für F_1 (Generalfaktor) .99, für F_2 (Extremitätenfaktor), .95, und für F_3 (Energieeinsatzfaktor) .94. Interessant ist, dass sich eine so hohe Übereinstimmung der Faktorenstruktur durchgesetzt hat, obgleich sich die Varianzanteile der Faktoren beträchtlich unterscheiden. Die Varianzanteile von F_1 bis F_3 betragen bei den Wiener Sportlern 63.8%, 11.1%, 10.1% (zusammen 84.9%), bei den Olympiasportlern 47.5%, 17.5%, 11.2% (zusammen 76.0%). Plausibel ist, dass bei den relativ weniger streng ausgelesenen Wiener Sportlern der Varianzanteil des Gesamtleistungsniveaus größer ist als bei den streng ausgelesenen Olympia-Sportlern, was sich in einem stärkeren ‚g'-Faktor-Anteil (F_1) bei den Wiener Sportlern niederschlägt. Nicht oder nur wenig ändern sich dadurch die Varianzquellen für F_2 und F_3, offenbar weil die anatomisch-physiologischen Leistungsbedingungen des Zehnkampfsports, die der Varianz von F_2 und F_3 zugrunde liegen, bei Spitzensportlern und den weniger Leistungsfähigen gleich ausgeprägt sind.

Übereinstimmung zwischen der Initial- und Variminlösung kann also bei F_2 und F_3 nicht die Rede sein. Varimin-F_2 ist sogar dem Initial-Faktor F_1 mit .66 etwas ähnlicher als dem Initial-Faktor F_2. Auch ist der Faktor Varimin-F_3 dem Faktor Initial-F_2 mit -.67 ähnlicher (mit verändertem Vorzeichen) als Initial-F_3.

Tabelle 4.06

Kongruenzen zwischen den Varimin-, Varimax- und Initialfaktoren
beim Datensatz Zarnowski (Olympische Leistungen N = 233)
Anmerkung: Kongruenzwerte ab .70 sind durch Fettdruck hervorgehoben

		a Varimin					b Varimin					c Varimax		
		F_1	F_2	F_3			F_1	F_2	F_3			F_1	F_2	F_3
	F_1	.86	.68	-.24		F_1	.98	.66	.28		F_1	.87	.62	.93
Varimax	F_2	.74	-.17	.41	Initial	F_2	.11	-.62	.67	Initial	F_2	-.38	.70	.06
	F_3	.86	.73	.55		F_3	-.15	.41	.69		F_3	-.32	-.34	.37

Was die Deutbarkeit von Initial-F_2 angeht, so zeigt Tabelle 4.05 beim 1500 m-Lauf zwar ebenfalls eine extrem hohe negative Ladung (-.76), eine Deutung in Richtung Ausdauer scheint also auch hier nahe zu liegen, zumal auch der Kugelstoß und Diskuswurf wie in der Varimin-Lösung mit hohen positiven Ladungen (.46 und .45) das Gegenteil von Ausdauer, d. h. Schnellkraft, repräsentieren. Doch fehlt bei Initial-F_2 eine hohe (Schnellkraft-) Ladung für den 100 m-Lauf. Der 1500 m- und der 100 m-Lauf sollten hinsichtlich der ausdauernden versus geballten Energieverausgabung einen idealen *minimal pair*-Vergleich ermöglichen. Doch bilden die beiden Laufsportarten in der Initial-Lösung kein minimales Paar, erst nach einer Varimin-Rotation lässt sich es finden. Ebenso fehlen bei Initial-F_2 Anzeichen für eine geballte Läuferleistung bei 110 m-Hürden, die Varimin aus den Korrelationen extrahiert.

Initial-F_3 wäre nicht ansatzweise als ein Faktor zu deuten, der Armkraft und Beinkraft differenziert, denn der Speerwurf (F_3 = -.45) sollte keine höhere Armkraft-Ladung auf Initial-F_3 haben als der Kugelstoß (-.22), und auch der 1500 m Lauf (-.40) sollte weniger Armkraft-Ladung haben als der Kugelstoß, was man bei einer Deutung von F_3 als Armkraft-Beinkraft-Polarität erwarten müsste. Auch ist eine Deutung von Initial-F_3 unter Heranziehung anderer Merkmale (das wäre ein denkbarer Ausweg) nicht in Sicht. Die Initialfaktoren F_2 und F_3 sind also schwer deutbar und viel weniger plausibel als die Variminfaktoren F_2 und F_3.

Zur Tabelle 4.06, die die Kongruenzwerte der faktoriellen Lösungsmöglichkeiten wiedergibt, bieten sich weitere Kommentare an:

1. Tabelle 4.06a: Der Varimin-Faktor F_1 zeigt relativ hohe Kongruenzen mit den drei Varimax-Faktoren (F_1, F_2, F_3). Dies ist offenbar auf die SS-Transformation zurückzuführen, die den Initialfaktor F_1, der mit Varimin konserviert bleibt bzw. optimiert wird (s. Tabelle 4.06), auf die drei Varimax-rotierten Faktoren ziemlich gleichmäßig verteilt.
2. Noch Tabelle 4.06a: Der Varimin-Faktor F_2 scheint mit Varimax-F_1 (.68) und -F_3 (.73) zusammenzuhängen, so als ob die Informationen des latenten Parameters, die in Varimin-F_2 bipolar zum Ausdruck kommen, sich auf Varimax-F_1 und -F_3 verteilt hätten.
3. Tabelle 4.06b: Abgesehen von der schon erwähnten Fast-Identität von Varimin-F_1 und Initial-F_1 sind kaum nennenswerte Übereinstimmungen zwischen der Varimin- und Initiallösung vorhanden. Dies wurde schon beim Deutungsversuch der Initialfaktoren auffällig.
4. Tabelle 4.06c: Drei erhöhte Kongruenzwerte zeigen sich beim Vergleich der Varimax- und Initialfaktoren, Varimax-F_3 liegt erstaunlich nahe (.93) bei Initial-F_1. Worauf diese Ähnlichkeiten beruhen, lässt sich schwer sagen, ist aber auch kaum untersuchenswert.

4. Expertenurteile zur Validierung der Varimin-Faktoren

Zwölf von 25 Zehnkämpfern, die übers Internet angesprochen wurden, beteiligten sich an der Aufgabe, den zehn Sportarten ihrer Disziplin Rangplätze zuzuweisen. 1. Rangordnung: *„Erster Rangplatz die Sportart, bei der es am meisten auf den Einsatz der Arme und am wenigsten auf den Einsatz der Beine ankommt, letzter Rangplatz die Sportart, bei der es am meisten auf den Einsatz der Beine und am wenigsten auf den der Arme ankommt"*. 2. Rangordnung: *„Erster Rangplatz die Sportart, bei der die Verausgabung der körperlichen Energie am meisten auf den Augenblick konzentriert wird (Schnellkraft), letzter Rangplatz die Sportart, bei der die Energieverausgabung über einen längeren Zeitraum zu erfolgen hat (Dauerleistung)."* Das Ziel der Befragung war es festzustellen, ob die Erfahrungen der Sportler bei der Ausübung der Sportarten mit den Deutungen der Variminfaktoren F_2 und F_3, die für die Rangordnungsaufgabe zu Kriterien erhoben wurden, in Übereinstimmung befinden. Abbildung 4.01 gibt auf der Y-Achse die mittleren Rangplätze der ersten Rangordnung wieder. Auf der X-Achse werden die Varimin-Faktorladungen abgetragen. Die Korrelation beträgt $r = .80$ ($p = .003$).

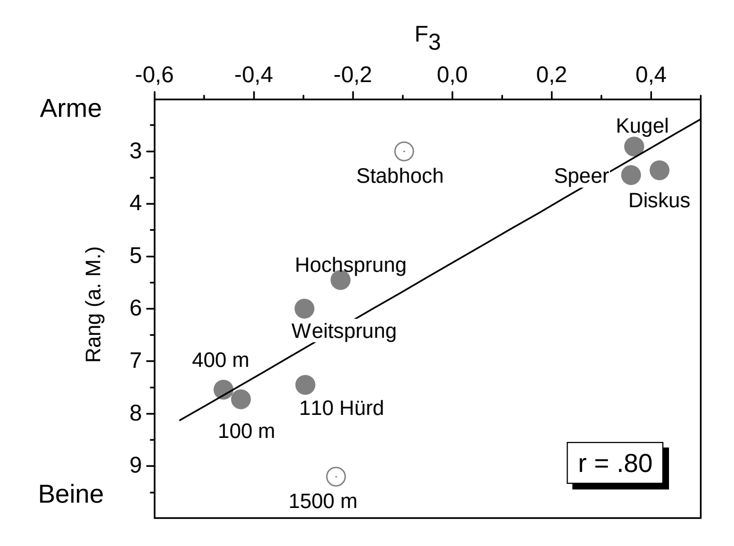


Abbildung 4.01
Die zehn Sportarten des Zehnkampfes mit Varimin-Faktorgewichten (X-Achse) sowie mittlere Rangplätze nach der Einstufung durch Zehnkämpfer hinsichtlich des relativen Arm-vs. Beineinsatzes (Y-Achse).

Es liegt nahe anzunehmen, dass die Sportler beim Stabhochsprung den Einsatz ihrer Arme, beim 1500 m Lauf den Einsatz ihrer Beine etwas überschätzt haben. Beim Stabhochsprung werden nach Kunz (1980, p. 166) die Funktionen des gesamten Körpers gleichanteilig gefordert (s. o., „Vielseitigkeit dieser Disziplin“). Beim 1500 m Lauf kommt es mehr auf eine anhaltende optimale Dosierung der Energieverausgabung als auf Beinkraft an. Dies wird mit Abbildung 4.02 plausibel, die die Ergebnisse der zweiten Rangordnung wiedergibt. Die Korrelation von r = .70 (p = .01) fällt geringer aus, offenbar vor allem weil die Sportler für den 400 m Lauf die erforderliche Ausdauer überschätzt und für das Speerwerfen unterschätzt haben. Denkbar wäre natürlich auch, dass nicht die Rangordnungen der Sportler, sondern die Faktorgewichte der Varimin-Analyse die Realität etwas verfälschen. Im Augenblick genügt es festzustellen, dass die Rangordnungen der Sportler insgesamt die Deutung der Variminfaktoren hinreichend bestätigen.

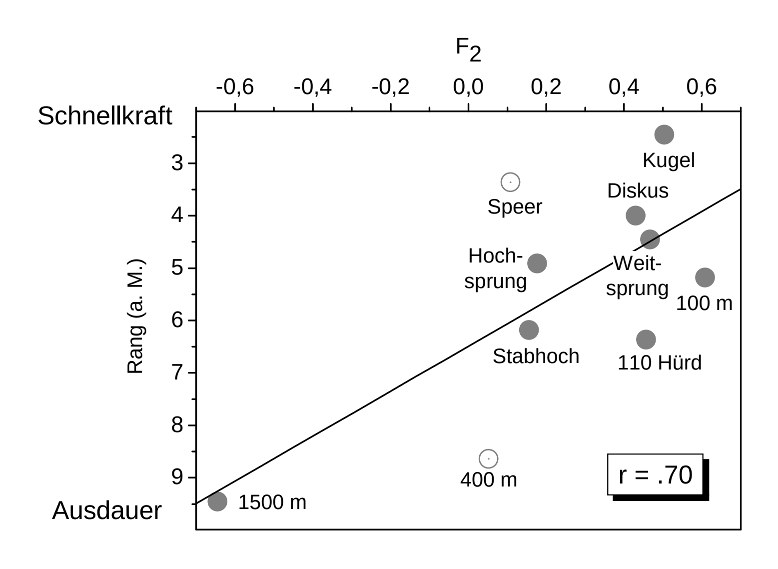


Abbildung 4.02

Die zehn Sportarten des Zehnkampfes mit Varimin-F_2-Gewichten (X-Achse) sowie mittlere Rangplätze nach einer Einstufung durch Zehnkämpfer hinsichtlich des relativen zeitlich konzentrierten vs. gestreckten Energieeinsatzes (Y-Achse).

Diskussion zu Kapitel 4

Die faktoriellen Reanalysen der sportlichen Leistungen machten deutlich, dass das konventionelle EFA-Paradigma, das dem *Simple Structure Modeling* (SSM) zugrunde liegt, im Ergebnis zu wünschen übrig lässt. Demgegenüber lieferte CSM (*Complex Structure Modeling*) transparente und leicht deutbare Strukturen. Auch waren die Variminfaktoren leichter deutbar als die Initialfaktoren. Zudem erwies sich die Variminlösung gegenüber der Varimaxlösung als stabiler. Darüber hinaus wurde die Deutung der Variminfaktoren durch Übereinstimmung mit den Selbstbeobachtungen erfahrener Zehnkampfsportler gestützt.

Die Varimin-Ergebnisse der vorliegenden Untersuchung stimmen zudem mit Ergebnissen sportpsychologischer Untersuchungen anderer Provenienz überein, so etwa mit Ergebnissen von Szopa et al. (1998), der multifunktionale Zusammenhänge suchte, ohne die Faktorenanalyse heranzuziehen. Die Forscher verwendeten 42 Tests motorischer Leistungen bei 143 männlichen und 91 weiblichen

Teilnehmern. Sie konnten fünf motorische Fähigkeiten unterscheiden, die drei ersten und wichtigsten stimmen mit den hier ermittelten Variminfaktoren überein. Szopa et al.'s „ability to develop global strength" manifestierte sich bei uns als Varimin F_1-Faktor, die „ability to develop local strength (of lower or upper extremities)" wurde bei uns als Varimin F_3-Faktor manifest, und die „ability of muscular endurance" entspricht einer polaren Ausprägung des bipolaren Varimin F_2-Faktors. Zwei weitere „abilities", die diese Autoren nennen, sind spezifisch-physiologisch definiert, sie sprechen von der Bildung von „alactacid and lactacid" in aerobischen (sauerstofferreich ablaufenden) Tests, und von maximalem „oxygen uptake" etc., was hier vernachlässigt werden kann.

Die Bipolarität der Energieverausgabung (*explosive* vs. *enduring*), wie sie durch Varimin-F_2 zutage trat, wird durch zusätzliche physiologische Beobachtungen bekräftigt: *"The energy at muscular activity can be supplied either a) anaerobically, as it is during short bursts of activity of high intensity with an accumulation of lactic acid as a result, or b) aerobically, as during more prolonged work, when oxygen intake balances the oxygen demand….In aerobic work, respiration and circulation will play a dominant role… "* (Astrand, 1956, p. 307). Die physiologischen Funktionen (a) und (b) sind trotz bipolarer Steigerungsmöglichkeit relativ unabhängig voneinander einsetzbar: *"The development of the anaerobic and aerobic processes are not parallel. It cannot, therefore be expected that a test procedure where the capacity of the aerobic processes is determined will give accurate information about the capacity of the man for aerobic work"* (Astrand, 1956, p. 307). Nach Milhorn H. T., 1982, ist als physiologische Grundbedingung für körperliche Durchhaltefähigkeit die Fitness des kardiovaskulären Systems anzusehen.

Die Anforderungen, die durch die sportliche Aufgabe gestellt werden, beanspruchen selten die verschiedenen Muskelpartien des Körpers in gleichem Maße. Von daher wird der Variminfaktor F_3 verständlich, der zeigt, dass entweder der oberen oder der unteren Partie des motorischen Körpersystems beim Einsatz Priorität zukommen kann. Experten werden durch dieses Ergebnis kaum überrascht sein. Sportler, die sich auf verschiedenen Gebieten bewähren wollen, *"take the challenge of competing across the whole range of athletic disciplines…" [but] "it is not possible for a driving mechanism to exist which is equally optimal or which can provide an identical maximum neural activation for all disciplines"* (Tidow, 2000, p. 6). Bestimmte „movement affinities" kommen vor, verschiedene Tätigkeiten erfordern den Einsatz verschiedener Muskel-Teilsysteme.

Die Ergebnisse mit V*arimax* -Transformation lassen sich nicht mit physiologischen Erkenntnissen aus anderen Quellen verbinden. Die SSM-Transformationen von Faktoren der Zehnkampfvariablen waren enttäuschend. Manning (1987) fasst die kläglichen Befunde, die bei konventioneller Orientierung mit anaerobischen Tests erzielt wurden, prägnant zusammen: *"Results showed no single factor emerged and that unrelated aspects existed among these tests and that they were not measuring similar quali-*

ties. It is suggested that anaerobic tests that are used to evaluate anaerobic power be performed as specifically as the skill being tested" – mit anderen Worten: Die Faktorenanalyse der konventionellen Art ist für eine Analyse sportlicher Leistungen unergiebig, auf sie sollte man verzichten, schlussfolgert der Autor. Warum will man bei körperlichen Leistungen des Menschen auf Faktorenanalysen gänzlich verzichten, nicht aber bei mentalen Leistungen und Dispositionen? Offenbar weil man das, was man an Ergebnissen zu erwarten hat, bei den vieldeutigen mentalen Prozessen und Dispositionen nicht so leicht prüfen kann.

Die komplexitätsoffene Alternative der faktorenanalytischen Forschung lässt weitere nützliche Ergebnisse von sportpsychologischen und -physiologischen Untersuchungen erwarten. Das Ziel der Forscherpioniere früherer Generationen (Guilford, 1958, Pöhlmann et al., 1979, Bös & Mechling, 1984 etc.) ist näher gerückt. Auf Thurstones erkenntnishemmendes SS-Prinzip sollte man bei der Erforschung von Komponenten der körperlichen Fitness (Ismail et al., 1965) verzichten. So hoffnungslos wie scharfe Kritiker die Situation der exploratorischen faktorenanalytischen Forschung sie darstellten (Lykken, 1991, Mitchell, 1997, Koch, 1999, Breiman, 2001, Gigerenzer, 2004, Barrett, 2005), braucht sie nach einer Wende zur Offenheit gegenüber Komplexität nicht zu bleiben.[97]

---

[97] Ein Beispiel für herbe Kritik am Versagen der konventionellen multivariaten Forschung sei einem Text von Barrett entnommen (Barrett, 2005, p. 45): *"…The pressure for change is building – and it looks like a paradigm change – for example, not merely a transition from say Classical Test Theory to IRT – but the entire loss …of psychometric test theory altogether over time."* Borsboom's "attack of the psychometricians", auf den sich Barrett bezieht, wird als ein Versuch betrachtet, die „selbst-verursachte Kluft zwischen der Psychometrie und der Psychologie" zu überbrücken. Das Modellieren der Psychometrie, ein Modellieren ohne Substanz, habe einem Modellieren mit Substanz zu weichen (Borsboom, 2006).

Einsichten und Aussichten

Meine Auseinandersetzung mit den Irrtümern der faktorenanalytischen Forschung mag ‚rebellisch' erscheinen. Hätte ich etwas tun sollen, um diesen Eindruck zu vermeiden?

Ich meine, während meines ersten Querdenkens und in den anschließenden zwölf Jahren empirischer Prüfungen, von deren Ergebnissen ich nur einen Teil vorführen konnte, meiner Verantwortung bewusst gewesen zu sein. Ich habe versucht, meine These vom Irrtum eines Modellierens mit Einfachstrukturen mithilfe nachprüfbarer Daten zu belegen. Gleichzeitig habe ich Evidenzen gesucht und vorgelegt für die Vorzüge eines Modellierens mit Komplexstrukturen mithilfe der dafür notwendigen CSM-Methodik.

Es wäre mir nicht möglich zu behaupten, ich hätte das Ziel verfehlt. Ich kann auch nicht abschwächend sagen, bei der Variminrotation handele es sich um ein Rotationsverfahren, das mit pluralistischer Toleranz der langen Liste früher entwickelter Verfahren hinzuzufügen sei, denn diese alle sind SSM-verhaftet. Es ist mir auch nicht möglich anzunehmen, dass neuere Datenanalysen der SSM, die der Strukturgleichungsmodelle eingeschlossen, eine günstigere Zukunft hätten.

Stattdessen glaube ich sagen zu dürfen: Mit dem SS-Prinzip hat Thurstone und haben seine Nachfolger Weichen gestellt mit außerordentlich nachteiligen Konsequenzen bis zum heutigen Tag (2011). Durch das Umstellen der Weichen in die Gegenrichtung lassen sich die Hauptfehler beheben. Kaum zu vermeiden aber wird sein, dass dadurch der Widerstand skeptischer Konventionalisten geweckt wird. Besser wäre, wenn sich durch meine Initiative Skeptiker zu empirischen Anstrengungen herausgefordert sähen und ihre Konvention gegen meine Irrtumsbehauptung mit geeigneten Daten zu verteidigen suchten.

Neutrale Beobachter könnten ohne Parteinahme beides für möglich halten: Entweder am Ende der Auseinandersetzungen ein Scheitern des Querdenkers Suitbert Ertel, der sich den Mehrheitsauffassungen der Experten nicht fügen wollte und dafür abgestraft wird. Sie könnten es auch für möglich halten, dass nach vergeblicher Verteidigung des Prinzips der mathematischen Einfachheit am Ende die komplexstrukturelle Analyse empirisch obsiegt.

Mit der Gefahr zu übertreiben, die ein Innovator meist unterschätzt, wage ich vorauszusagen, dass eine komplexanalytische Auffassung der psychischen Phänomene generell einen neuen Zugang zu den latenten Prozessen eröffnen müsste, die ihnen zugrunde liegen. Das Chaos unanalysierter Konstrukte sollte durch ein komponenten-orientiertes Modellieren verschwinden. Neue Begriffe sollten auf dieser Route sichtbar werden und konstruktive Entwicklungen nach sich ziehen. Nach Überwindung der SSM-Misere sollte eine Wende zur CSM-Strategie das Gebot zukünftiger Forschung werden.

Was steht der Wende zur CSM hauptsächlich entgegen? Es ist die durch das scheinbar unverzichtbare SS-Ideal geschaffene Vorstellung, dass Wörter und Sätze, wenn man ihre Verwendungshäufigkeiten ermittelt, objektiviert und damit einer Bedeutungsanalyse enthoben werden, die sich nicht mehr auf etwas Gezähltes stützt. Unter dem SS-Einfluss werden begrifflich fassbare Variablen generell als unauflösbare Einheiten aufgefasst, seien sie manifest (mithilfe singulärer ‚Dimensionen' quantifizierbar) oder latent (als ‚Dimensionen' scheinbar nicht weiter erklärungsbedürftig). Man glaubt, ein Endziel erreicht zu haben, wenn einem Faktor der Rang einer ‚Dimension' verliehen wurde, man sucht nicht nach funktionalen Kontexten und Komponenten, aus denen heraus sich ein Dimensionskonstrukt begründen müsste.

Mathematische Operationen, die an den Daten vollzogen werden, sind mitunter unangemessen. Wenn man sich an ihre Ergebnisse klammert, auch wenn sie unangemessen sind, schränken sie die Freiheit des Problemlösens erheblich ein. Fundamentale Folgefehler werden nicht mehr erkannt. Konsequenz: Das Urteil über die Vorteile, die dem Forscher durch eine statistische Methode gegeben sind – es gibt solche reichlich natürlich – erfordert Einsicht. Diese aber ist nicht allein

aus der Methode herleitbar. Der höherwertige Rang der Vernunft sollte gegenüber dem Einfluss einer attraktiven und subjektive Sicherheit gebenden Mathematisierung nicht leichtfertig unterschätzt werden.

Zusammenfassungen der Kapitel

Kapitel 1: Kritik am Dogma „Simple Structure“

Die Studie hat zum Ziel, einen methodologischen Irrtum aufzudecken, der die faktorenanalytische Forschung seit ihren Anfängen belastet. Das Modell der *Simple Structure* (Thurstone, 1935/1947), an dem sich diese Forschung orientiert hat, wird infrage gestellt. Es wird aufgezeigt, dass durch die Rotation der Faktoren zur Einfachstruktur die Komplexität der Bedingungen des Erlebens und Verhaltens verschleiert wird. Circumplex- und Strukturgleichungs-Verfahren lösen die grundlegenden Probleme nicht, da sie die Unzulänglichkeiten der Einfachstruktur mit übernehmen. Die Kritik an Simple Structure wird am empirischen Beispiel erläutert, theoretisch begründet und mit zahlreichen wörtlichen Zitaten einer Minorität kritischer Autoren gestützt. Insgesamt wird die Kalamität der konventionellen Faktorenanalyse einem einseitig mathematischen Denken verantwortlicher Methodiker zur Last gelegt, das andere legitime Erkenntnismittel vernachlässigt. Ein Paradigmenwechsel wird für längst fällig gehalten. Als Alternative zur konventionellen Faktoren-Transformation wird eine Transformation zur Komplexstruktur gefordert, deren Ergebnisse beispielhaft in Kapitel 2 vorgestellt werden.

Kapitel 2: **Komplexe Strukturen aufspüren.** Faktorenanalysen mit Variminrotation.

Kapitel 2 knüpft an der Kritik am ‚Dogma Simple Structure' an, die in Kapitel 1 ausgeführt wird. Ein neues Rotationsverfahren für extrahierte Faktoren wird vorgestellt (Varimin), welches die latenten Kovarianzquellen ermitteln will, die den manifesten Variablen einer Domäne zugrunde liegen. Varimin optimiert das durch die Initialstruktur einer Faktorenanalyse schon entworfene, aber oft noch verbesserungsbedürftige Modell multifunktionaler Bedingungsverhältnisse. Die konventionell verwendeten Rotationsverfahren, die latente Kovarianzquellen als Einfachstruktur modellieren wollen, voran das Standardverfahren Varimax, werden bemängelt, weil sie die reale Komplexität der Bedingungsverhältnisse verdecken, anstatt sie aufzuklären. Die neue methodologische Ausrichtung wirft Fragen auf, von denen fünf diskutiert werden. Gefragt wird: Wie lassen sich Varimin-transformierte Faktoren deuten? Wird die Komplexität der Kovarianzquellen mit der Initiallösung nicht schon hinreichend repräsentiert? Lassen sich faktorielle Simple Structure-Lösungen nicht doch befriedigend interpretieren, wie könnte man sie sonst ständig verwenden? Enthält die regelmäßig anzutreffende Bipolarität von Varimin-Faktoren substantielle Information? Lassen sich mithilfe einer Komplexstruktur-orientierten Faktorenanalyse auch methodenbedingte Faktoren entdecken? Die Klärung dieser Fragen wird durch Anwendung der Varimin-Transformation auf acht publizierte und zwei neu generierte Datensätze unterstützt. Die Transformation zur Komplexstruktur, mit welcher latente Kovarianzquellen der manifesten Variablen aufgespürt werden (durch Varimin), wird der Transformation zur Einfachstruktur (durch Varimax) gegenüber gestellt. Es zeigt sich, dass Varimax zur Aufklärung latenter Bedingungen keinen Beitrag leistet. Lediglich zum Clustern manifester Variablen sind Simple Structure-Transformationen geeignet. Auf die Grenzen der methodologischen Innovation Varimin wird hingewiesen.

Kapitel 3: **Komplexität intellektueller Leistungen**: Grundintelligenz *(g)* und Lernkapital *(l)*.

Achtzehn der Forschungsliteratur entnommene Interkorrelationstabellen des Intelligenz-Struktur-Tests (I-S-T) (Versionen 1955 und 1970) wurden PCA-faktorisiert und durch Varimin-Rotationen einer komplexstrukturellen Modellierung zugeführt. Die achtzehn sich ergebenden Varimin-Lösungen wurden aggregiert, es resultierten zwei Faktoren: Varimin-F_1 war als Generalfaktor *g* zu deuten, Varimin-F_2 ließ eine von *g* unabhängige leistungsmodifizierende Varianzquelle zutage treten. Sie wurde als Auswirkung eines durch vorhergehendes schulisches und durch anderes Training erworbenen *Lernkapitals* (*l*) gedeutet, von welchem die

Subtestleistungen des I-S-T mehr oder weniger profitieren. Die Validität von Varimin-F_1 als *Grundintelligenz g* wurde durch hohe Korrelationen zwischen *g* und Testscores der allgemeinen Intelligenz erhärtet (operationalisiert durch CFT und FRT). Auch wurde die Deutung von Varimin-F_2 als Lernkapital bekräftigt, da F_2-Faktorscores mit den trainingabhängigen Schulnoten, mit der Leistung im Rechtschreiben und im Grundrechnen signifikant korrelierten. Die achtzehn PCA-faktorisierten Interkorrelationstabellen wurden auch einer simple structure-Modellierung zugeführt (durch Varimax-Rotation). Varimax-F_1 entsprach inhaltlich dem, was man ‚fluide' Intelligenz zu nennen pflegt, Varimax-F_2 entsprach der konventionell so genannten ‚kristallinen' Intelligenz. Doch fehlten zwischen Varimax-F_1 und F_2 zu erwartende Unterschiede ihrer Korrelationen mit den Außenkriterien der allgemeinen Intelligenz, der Schulnoten, des Rechtschreibens und des Grundrechnens, die ihre Bedeutung als ‚fluide' vs. ‚kristalline' Intelligenz manifestiert haben sollten. Zudem hatte das Aggregat der Varimax-Faktoren einen dritten substantiellen Faktor ergeben, der nicht interpretierbar ist und offensichtlich ein Kunstprodukt darstellt. Durch die Varimax-Rotation wird initiale *g*-Varianz auf F_1, F_2 und, wenn vorhanden, auch auf F_3 künstlich aufgeteilt. Dabei wird der Varianzbeitrag des Lernvorteils mit dem der Grundintelligenz verschmolzen. Die Einfachstruktur-Modellierung gibt den latenten Funktionskomponenten (*g*) und (*l*) keine Chance, sich unabhängig voneinander zu manifestieren.

Kapitel 4: **Komplexität sportlicher Leistungen**. Generelle sportliche Kondition, Ort des Energieeinsatzes, Dosierung des Energieeinsatzes.

Mithilfe von Zehnkampfdaten aus vier Quellen wurde die Effizienz der Varimin-Faktorenrotation überprüft. Es zeigte sich, dass sich die Varimin-Faktoren als Manifestation latenter Varianzquellen deuten ließen, die den sportlichen Leistungen multifunktional zugrunde liegen. Varimax-Faktoren, die das Modell der Einfachstruktur zu optimieren suchen, führten zu fragwürdigen Clustern der manifesten Variablen. Die Varimin-rotierten Faktorenstrukturen erwiesen sich gegenüber einem Wechsel der Datenquellen auch als robuster als die Varimax-rotierten Strukturen. Eine Anwendung der exploratorischen Faktorenanalyse, die auf Erfassung von Komplexität ausgerichtet ist, könnte die in Misskredit geratene exploratorische Forschungsstrategie rehabilitieren.

Abstracts of Chapters

Chapter 1: **Criticism of Thurstone's "Simple Structure"**.

The present study aims at disclosing a methodological prejudice which has misled factor analysis since its beginning. Simple structure, the guiding principle for factor rotation (Thurstone, 1935/1947), is regarded as a questionable dogma. The Thurstone principle impedes the discovery of latent sources of variance on which manifest empirical variables are based. This criticism is elucidated by theoretical considerations and by verbatim quotations of critical authors. The present calamity of factorial research is deemed to be due to general flaws of methodical reasoning. One-sided mathematical formalization in the discipline has lost its objectives by ignoring ordinary sources of knowledge and linguistic processing. The problem of simple structure cannot be solved by circumplex- and structural equation procedures which suffer from the same flaws as simple structure modelling. A paradigm change is overdue. An alternative factor transformation modeling complex structures is demanded, it will be delineated in chapter 2.

Chapter 2: **Unraveling complex structure**. Factor analyses with varimin rotation.

This chapter is a sequel of chapter 1 (Ertel, 2009) in which varimax, the procedure of factor rotation commonly used to generate simple structure, is replaced with varimin. Varimin aims at manifesting interacting aspects of latent structural components. The model of complexity which is outlined by initial factor extractions is optimized. The new method raises a number of issues of which five are discussed at length. (1) Are varimin factors interpretable? (2) Can latent sources of covariance, being complex already by initial solutions, be rendered more complex? (3) Have simple structure solutions in common practice not been tolerable, to say the least? (4) Varimin factor loadings are often bipolar. How to interpret the bipolarity of varimin factors? (5) Does varimin reveal contributions of variance originating from methodical sources? Ten empirical applications of varimin transformation serve as examples. Particular features of transformation to complex structure, revealing latent sources of covariance (by varimin), are compared with pertinent results obtained from transformations to simple structure (by varimax). Varimax will remain useful merely for clustering objectives. Attention is also drawn to limitations of the methodical innovation.

Chapter 3: Complexity of intellectual performance. Basic intelligence *(g)* und learning assets *(l)*.

Eighteen matrices with intercorrelations of eight subtest variables of the intelligence test I-S-T, widely used in Germany, were subjected to principal component analysis, the resulting factors were rotated by varimin towards optimal complex structure. The 18 varimin solutions were aggregated, two factors resulted: Varimin-F_1 represented a general factor *g* ('general or basic intelligence'), varimin-F_2 represented an achievement-modifying factor, obviously dependent on preceding educational training and learning ('learning assets', *l*). The validity of varimin-F_1, general intelligence, was ascertained by high correlations between g and test scores of general intelligence, operationalized by culture-free tests CFT und FRT. The interpretation of varimin-F_2 found support by significant correlations with learning-dependent school grades and scores in orthography and arithmetic. The 18 PCA-factors were also transformed by varimax to simple structure. This transformation caused a splitting up of initial *g* into two seemingly separate factors, called 'fluid' and 'crystallized' intelligence by convention. In addition, differences between varimax F_1 (fluid) and F_2 (crystallized) regarding correlations with external criteria (general intelligence vs. school grades, training scores in orthography and arithmetic) which should have emerged were missing. The aggregate of varimax results yielded an unexpected third factor with considerable weight which, however, turned out to be an artifact of simple structure. Apparently, simple

structure modelling of intelligence test data blends general intelligence with learning effects. Rotation of intelligence data to simple structure does not reveal, as it should, independent contributions of latent functional components.

Chapter 4: Complexity of physical performance: Overall physical condition, locus of energy expenditure, dosage of energy expenditure.

The results of a factorial study are reported using sports data whose underlying sources of variance are more transparent than those of invisible mental data: decathlon record scores. The aim was to compare varimin and varimax results regarding factorial stability and interpretability. It is shown that varimin factors revealed latent sources of variance in interaction, while varimax factors yielded obscure clusters of those sources of variance. In addition, factor structures obtained by varimin rotation were more robust to changing data sources than those obtained by varimax rotation. The new methodical turn might revive exploratory factorial research which, due to questionable results of the past, has lost reputation.

Literatur

Acton, G. S. & Revelle, W. (2004). Evaluation of the psychometric criteria for circumplex structure. *Methods of Psychological Research Online, 9 (1),* 1-29.

Amthauer, R. (1970). IST70. *Intelligenz-Struktur-Test 70*. Göttingen: Hogrefe.

Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (1999). *I-S-T-2000. Intelligenz-Struktur-Test 2000*. Göttingen: Hogrefe.

Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *I-S-T-2000-R. Intelligenz-Struktur-Test 2000-R*. Göttingen: Hogrefe.

Andresen, B. (1998). *Persönlichkeitsstrukur und Psychosetendenz. Bd. 1: Grundlagen, Methoden, Basisfaktoren.*Westerau: Psychometrie- & Psychodiagnostik-Verlag.

Arkes, H. R. (1991). Costs and benefits of judgment errors: implications for debiasing. *Psychological Bulletin, 110 (33),* 486-498.

Astrand, P.-O. (1956). Human physical fitness with special reference to sex and age. *Physiological Reviews, 36,* 307-329.

Barrett, P. (2002). Beyond psychometrics: Measurement, non-quantitative structure, and applied numerics. *Augmented web version from The Journal of Managerial Psychology.*

Barrett,P.(2005). Validity and utility in I/O psychology. Retrievable from the web. www.pbarrett.net.

Basilevski, A. (1994). *Statistical factor analysis and related methods. Theory and applications.* New York: Wiley.

Bates, E., Thal, D., & Janowsky, J. (1992). Early language development and its neural correlates. In I. Rapin & S. Segalowitz (Eds.), *Handbook of neuropsychology*, Vol. 7: Child neuropsychology. Amsterdam: Elsevier.

Bäumler, G. & Breitenbach, W. (1970). Zusammenhänge zwischen Intelligenz, Konzentration, Angst und Leistungsmotivation bei einer studentischen Stichprobe. *Psychologie und Praxis*, *14*, 37-40.

Bäumler, G. & Weiß, R. (1966). Über den Zusammenhang der Paulitestleistung mit Intelligenzleistungen (IST-Amthauer, CFT-Cattell). *Psychologie und* Praxis, 10,27-36.

Beauducel, A. & Wittmann, W.W. (2005). Simulation study on fit indices in confirmatory factor analysis based on data with slightly distorted simple structure. *Structural Equation Modeling, 12,* 41-75.

Beauducel, A., Strobel, A., & Brocke, B. (2003). Psychometrische Eigenschaften und Normen einer deutschsprachigen Fassung der Sensation Seeking-Skalen, Form V. *Diagnostica, 49(2),* 61-72.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity*. Princeton NJ: Princeton University Press.

Bem, S.L.(1981). Gender schema theory: A cognitive account of sex typing. *Psychological* Review, 88, 354-364.

Berneyer, G. (1957). Psychological factors. Their number, nature, and identification. *The British Journal of Statistical Psychology, 10*, 17-27.

Bierhoff, H. W. (2000). Skala der sozialen Verantwortung nach Berkowitz und Daniels: Entwicklung und Validierung. *Diagnostica, 46 (1),* 18-28.

Blanton, H. & Jaccard, J. (2006) Arbitrary metrics in psychology. *American Psychologist*, *61* 27-41.

Blinkhorn, S. F. (1997). Past imperfect, future conditional: Fifty years of test theory. *British Journal of Mathematical and Statistical Psychology, 50*, 175-185.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117* (*2*), 187-215.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrica*, *71*, 425-440.

Bortz, J. (2004). *Lehrbuch der Statistik*. Sechste Auflage. New York: Springer. Erste Auflage 1977.

Bös, K. & Mechling, H. (1984) Dimensionen sportmotorischer Leistungen als Grundlage für einen fähigkeitsorientierten Ansatz. In G. Hagedorn, H. Karl & K. Bös (Red.) *Handeln im Sport*. Sechster Sportwissenschaftlicher Hochschultag. Clausthal-Zellerfeld.

Bös, K. (1987) *Handbuch sportmotorischer Tests*. Göttingen: Hogrefe.

Breiman, L.(2001).Statistical modeling: The two cultures. *Statistical Science*, *16* (3), 199-215.

Bridgman, P.W. (1959). *The way things are*. Cambridge: Harvard University Press.

Briggs, S. R. (1989). The optimal level of measurement for personality constructs. In D. M. Buss & N. Cantor (Eds.), *Personality Psychology. Recent Trends and Emerging Directions* (pp. 246-260). New York: Springer.

Brocke, B., Beauducel, A., & Tasche, K. (1998). Der Intelligenz-Struktur-Test: Analysen zur theoretischen Grundlage und technischen Güte. *Diagnostica,* 44, 84-99.

Burroughs, G. E. R., & Miller, H. W. L. (1961). The rotation of principal components. *The British Journal of Statistical Psychology, 14 (Part 1)*, 35-42.

Burt, C. (1949). The structure of the mind: A review of the results of factor analysis. *The British Journal of Educational Psychology,19*, 100-111, 176-199.

Burt, C. (1954). The sign pattern of factor matrices. *The British Journal of Statistical Psychology, 7, Part I*, 15-29.

Büsch, D., Hagemann, N. & Thielke, S. (2001) Gibt es eine Ballfähigkeit. Interpretation eines faktorenanalytischen Ergebnisses. *Psychologie und Sport, 8* (2), 57-66.

Buse, L. & Pawlik, K. (1978). Stichprobenfehler und Rotationsambiguität von Faktorladungen. *Zeitschrift für experimentelle und angewandte Psychologie*, *25*, 10-34.

Butler, J. M. (1969). Simple structure reconsidered: Distinguishability and invariance in factor analysis. *Multivariate Behavioral Research*, *4*, 5.

Campbell, D. T. & Fiske, D. W. (1959), Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Carl, W. (1968). Eine Untersuchung zur Faktorenstruktur von Antworttendenzen (‚response sets') bei Antwortskalen verschiedener Stufenzahl. *Zeitschrift für experimentelle und angewandte Psychologie, 15*, 419-434.

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor analytic studies.* Cambridge: Cambridge University Press.

Caspi, A. (1998). *Social, Emotional, and Personality Development.* In W. Damon & N. Eisenberg (Eds.), *Handbook of Child Psychology* (Vol. *3*, pp. 311-388). New York: John Wiley & Sons.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence. A critical experiment. *Journal of Educational Psychology, 65,* 1-22.

Cattell, R. B. (1965). Factor analysis: An introduction to its essentials. II: The role of factor analysis in research. *Biometrics, 21,* 405-435.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* New York: Houghton Mifflin.

Cattell, R. B., & Dickman, K. (1962). A dynamic model of physical influences demonstrating the necessity of oblique simple structure. *Psychological Bulletin, 59*, 389-400.

Cattell, R. B., & Radcliffe, J. A. (1962). Reliabilities and validities of simple and extended weighted and buffered unifactor scales. *The British Journal of Statistical Psychology, 15*(*2*), 115-128.

Child, D. (1990). *The essentials of factor analysis*. London: Cassell, 1990, 2nd. Edition, 1st edition: 1971.

Church, A. T., & Burke, P. J. (1994). Exploratory and confirmatory tests of the big five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology, 60*, 93-114.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18*, 115-126.

Cliff, N. (1992). Abstract measurement theory and the revolution that has never happened. *Psychological Science*, *3,* 186-190.

Comrey, A. L. (1978). Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology, 46(4),* 648-659:

Comrey, L., & Lee, H. (1992). *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum, 2nd Edition. 1st Edition Comrey alone: 1973, New York: Academic Press.

Crawford, J. R. & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *43*, 245-265.

Cronemeyer, I. (1983). Analyse der Intelligenz- und Persönlichkeitsstruktur von Abiturienten der neu gestalteten Gymnasialen Oberstufe in Nordrhein-Westfalen. Eine Untersuchung über die Beziehungen zwischen der Abiturdurchschnittsnote und dem „Intelligenz-Struktur-Test" von Amthauer sowie dem „Freiburger-Persönlichkeits-Inventar". *Dissertation. Erziehungswissenschaftliche Fakultät Universität Köln.*

Davies, W. K. D. (1971). Varimax and the destruction of generality: a methodological note. *Area, 3*, 112-118.

De Raad, B., Hendriks, A. A. J., & Hofstee, W. K. B. (1994). The Big Five: a tip of the iceberg of individual differences. In C. J. Halverson, Jr., G. A. Kohnstamm, and R. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 91–109). Hillsdale, NJ: Erlbaum.

Diener, E- & Emmons, R. A. (1983). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, *47, 5,* 1105-1117.

Edmonds, B. (2002). *Simplicity is not truth-indicative* (*No. CPM-OL-99*).

Ertel, S. (2009a). Kritik am Dogma „Simple Structure“. *Psychologie des Alltagshandelns, 2(1),* 36-63.

Ertel, S. (2009b). Latente Strukturen aufspüren. *Faktorenanalysen mit Variminrotation. Psychologie des Alltagshandelns 2(2),* 33-60.

Ertel, S. (2010a). Grundintelligenz (*g*) und Lernkapital (*l*). Komplexstrukturelle Faktorenanalyse des I-S-T. *Psychologie des Alltagshandelns 3 (2),* 49-61.

Ertel, S. (2011). Exploratory factor analysis revealing complex structure Personality and Individual Differences. 50, 196-200

Ertel, S. (a) Varianzquellen der Persönlichkeit. Exploratorische Faktorenanalyse mit komplextheoretischer Ausrichtung. (*Unpublished*).

Ertel, S. (b) Faktorielle Komplexstruktur durch Varimin-Rotation. Manifestation simulierter Zusatzvarianz in I-S-T-Daten. (*Unpublished*).

Ertel, S. (c). Simulierte Extravarianz in I-S-T-Daten, manifestiert durch faktorielle Variminrotation. (Unpublished).

Estes, W. K. (1975). Some targets for Mathematical Psychology. *Journal of Mathematical Psychology*, *12,* 263-282.

Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and Individual Differences, 13*, 667-673.

Eysenck, H. J. (1997). Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology, 73*, 1224 - 1237.

Fabrigar, L. R., Wegener, D. Z., McCallum, R. C. & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299.

Feger, H. (1971). Einstellungsstruktur und Einstellungsänderung: Ergebnisse, Probleme und ein Komponentenmodell der Einstellungsobjekte. *Zeitschrift für Sozialpsychologie,* 10, 331-349.

Feger, H. & Brehm, M. (2001). *New developments in feature pattern analysis.* Lengerich: Pabst Science Publishers.

Ferrando, P. J. & Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: Some aspects of the problem and some suggestions. *Psicologica, 21* 301-323.

Finch, J. F., & West, S. G. (1997). The investigation of personality structure: Statistical models. *Journal of Research in Personality, 31*, 439-485.

Fischer, H. (1958). Ein Vergleich zwischen dem IST von Amthauer und dem PMA von Thurstone. *Diagnostica, 4,* 25-32.

Fittkau, B. (1968). Die Mehrdeutigkeit der Varimax-Rotation und einige sich daraus ergebende Folgerungen. *Archiv für die gesamte Psychologie, 120*, 106-114.

Fleck, L. (1935). *Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv.* Basel: Schwabe.

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology. 39*, 291-314.

Forster, M. (1998). Parsimony and simplicity. *Available from: http://philosophy.wisc.edu/forster/220/simplicity.html*.

Forster, M., & Sober, E. (1994). How to tell when simple, more unified, or less ad hoc theories will provide more accurate predictions. *The British Journal for the Philosophy of Science, 45*, 1-35.

Funder, D. C. (2001). Personality. *Annual Review of Psychology, 52*, 197-121.

Gangestad, S. & Snyder, M. (1985). "To carve nature at its joints": On the existence of discrete classes in personality. *Psychological Review, 92.* 317-349.

Gawronski, B. (2000). Falsifikationismus und Holismus in der experimentellen Psychologie: Logische Grundlagen und methodologische Konsequenzen. *Zeitschrift für Sozialpsychologie*, *31 (1),* 3-17.

Gibbons, R. D. & Hederer, D. R.(1992). Full-information item bi-factor analysis. *Psychometrika*, *57,* 423-436.

Gigerenzer, G. & Strube, G. (1978) Zur Revision der üblichen Anwendung dimensionsanalytischer Verfahren. *Zeitschrift für Entwicklungspsychologie und Pädagogischen Psychologie 10(1),* 75-86.

Gigerenzer, G. (1978). Artefakte in der dimensionsanalytischen Erfassung von Urteilsstrukturen. *Zeitschrift für Sozialpsychologie, 9*, 110-116.

Gigerenzer.G.(2004). Mindless statistics. *The Journal of Socio-Economics*, *33,* 587-606.

Glass, G. V. & Taylor, P. A. (1966). Factor analytic methodology. *Review of Educational Research*, *36,* 566-587.

Glass, G. V. (1991). Eulogy for Henry Kaiser (1927-1992). *Multivariate Behavioral Research Monographs, 27*, 159-171.

Goldberg, L. R. (1981). Developing a taxonomy of trait-descriptive terms. In D. W. Fiske (Ed.), *Problems with language. New directions for methodology of social and behavioral science, No. 6* (pp. 43-65). San Francisco: Jossey-Bass.

Goldberg, L. R. (1990). An alternative "Description of Personality": The Big-Five Factor Structure. *Journal of Personality and Social Psychology*, *59*, 1216-1229.

Gorsuch, R. L. (1983). *Factor analysis*. Philadelphia: Saunders. 1st Edition: 1974.

Gould, S. J. (2002). *The Structure of Evolutionary Theory*. Cambridge (MA): Harvard UP-Belknap Press.

Greif, S.(1972). *Gruppenintelligenztests –Untersuchungen am WIT, IST, LPS und AIT.* Frankfurt: Peter Lang.

Groner, R., & Groner, M. T. (1991). Heuristische versus algorithmische Orientierung als Dimension des individuellen kognitiven Stils. In K. Grawe, R. H. Hänni, N. Semmer & F. Tschan (Eds.), *Über die richtige Art, Psychologie zu betreiben*. Göttingen: Hogrefe.

Guilford, J. P. & Hoepfner,R. (1969).Comparison of varimax rotations with rotations to theoretical targets. *Educational and Psychological Measurement*, *29*, 3-23.

Guilford, J. P. (1952). When not to factor analyze. *Psychological Bulletin*, *49*, 26-37.

Guilford, J. P. (1958) A system of the psychomotor abilities. *American Journal of Psychology,71* (1), 164-174.

Guilford, J. P., & Zimmerman, W. S. (1963). Some variable-sampling problems in the rotation of axes in factor analysis. *Psychological Bulletin, 60*, 289-301.

Guilford, J.P. (1974). Rotation problems in factor analysis. *Psychological Bulletin, 81*, 4958-501.

Guttman, L. (1954). A new approach to factor analysis: The radix. In P. F. Lazarsfeld (Ed.) *Mathematical Thinking in the social sciences* (pp. 258-348).Glencoe, IL: Free Press.

Guttman, L. (1955). The determinacy of factor score matrices, with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology*, *8*, 65-81.

Guttman, L. (1971). Measurement as structural theory. *Psychometrica, 36*, 329-347.

Guttman, L. (1992). The irrelevance of factor analysis for the study of group differences. *Multivariate Behavioral Research*, *27,* 175-204.

Haenschke-Kramer, H. & Mehl, J. (1967). Zur Untersuchung von Spezialbefähigungen auf mathematisch-naturwissenschaftlichem Gebiet.
Zeitschrift für Psychologie,174, 285-310.

Halberg, F. (1980). Closing remarks: Chronobiology and health. In L. E. Scheving & F. Halberg (Eds.), *Chronobiology: Principles and applications to shifts in schedules* (pp. 541-562). Alphen aan den Rijn: Sijthoff and Noordhoff.

Hamilton, M. (1958). An experimental approach to the identification of factors. *The British Journal of Statistical Psychology, 11* (*2*), 161-169.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurosurgical Psychiatry, 23*, 56-62.

Harmann, H.H. (1967). *Modern factor analysis.* Chicago: The University of Chicago Press.

Herrmann, T. (1965). Psychologie der kognitiven Ordnung.. In C. F. Graumann & J. Linschoten (Eds.), *Phänomenologisch-Psychologische Forschungen.* (Vol. 6). Berlin: Walter de Gruyter & Co.

Highmore, G., & Taylor, W. R. (1954). A factorial analysis of athletic ability. *British Journal of Statistical Psychology, 7*, 1-8.

Hinz, A., Brähler, E., Geyer, M., & Körner, A. (2003). Urteilseffekte beim NEO-FFI. *Diagnostica, 49*, 157-163.

Hofstee, W. K. B., de Raad, B., & Goldberg, L. R. (1992). Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology, 63*, 147-163.

Höger, D.(1964). Analyse der Intelligenzstruktur bei männlichen Gymnasiasten der Klassen 6-9 (Untersekunda-Oberprima). *Psychologische Forschung, 27*, 419-474.

Holling, H., Preckel, F. & Vock, M. (2004). *Intelligenzdiagnostik. Kompendien Psychologische Diagnostik*. Band 6. Göttingen: Hogrefe.

Holz-Ebeling, F. (1995). Faktorenanalysen und was dann? *Psychologische Rundschau, 46*, 18-35.

Holzinger, K. J. & Swineford, F. (1939). A study of factor analysis: The staibilty of a bi-factor solution. *Supplementray Educational Monographs 48*. Chicago: University of Chicago, Department of Education.

Horn, J. L. (1967). On subjectivity in factor analysis. *Educational and Psychological Measurement, 27*, 811-820.

Horn, J. L. (1976). Human abilities: A review of research and theory in the early 1970's. *Annual Review of Psychology, 27,* 437-485.

Humphreys, L. G. & Davey, T. C. (1988). Continuity in intellctual growth from 12 months to 9 years. *Intelligence*, *12*, 183-197.

Ismail, A. H., Falls, H. B., MacLeod, D. F.(1965). Development of a criterion for physical fitness tests from factor analysis results. *Journal of Applied Physiology*, *20* (5), 991-999.

Jacobson, R. & Halle.(1956). *Fundamentals of Language*. Den Haag: Mouton.

Jäger, A. O., & Hörmann, H. (1981). Demonstrationen von „g“ (der Allgemeinen Intelligenz) und zur Bedeutung des Variablenkontextes bei exploratorischen Faktorenanalysen. *Psychologische Beiträge, 23*, 408-420.

Jensen, A. R. (1998). *The g factor. The science of mental ability.*Westpor: Praeger.

Johnson, W. & Bouchard, T. J. (2005. The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence,* 33, 393-416.

Jöreskog, K. G. & Sörbom, D. (2003) LISREL 8.52. *Website*: www.ssicentral.com .

Just M. A. & Carpenter, P. A. (1984). *The psychology of reading and language comprehension*. Boston: Allyn and Bacon.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187-200.

Kaiser, H. F. (1970). A second generation "Little Jiffy". *Psychometrika, 35*, 401-416.

Kallina, H. (1967). Das Unbehagen in der Faktorenanalyse. *Psychologische Beiträge, 10*, 81-86.

Karvonen, M. J. & Niemi, N.(1953). Factor analysis of performance in track and field events. *European Journal of Applied Physiology*, *15 (2),* 1439-6327. Identical publication in *Arbeitsphysiologie*, *15*, 127-133.

Koch, S.(1999). Psychology in human context. Essays in dissidence and reconstruction. Chicago: The university of Chicago Press.

Kallina, H. (1967). Das Unbehagen in der Faktorenanalyse. Sonderheft: Münchener Symposium über Faktorenanalye. *Psychologische Beiträge, 10*, 81-86.

Kalveram, K. T. (1970). Über Faktorenanalyse. Kritik eines theoretischen Konzepts und seine mathematische Neuformulierung. *Archiv für die Psychologie, 122*, 92-118.

Kanizsa, G. (1975). „Prägnanz" as an obstacle to problem solving. *Giornale Italiano do Psicologia, 2(3),* 417-425.

Kaplunovsky, A. S. (2007). Why using factor analysis? Dedicated to the centenary of factor analysis. HAIT Journal of Science and Engeneering. Prepublication draft, Internet-zugänglich.

Kaplunovsky, A. S. (2005). Factor Analysis in Environmental Studies, *HAIT Journal of Science and Engineering B 2, 1-2*, 54-94.

Kazden, A.E. (2006) Arbitrary metrics: Implications for identifying evidence-based treatments. *American Psychologist*, 61 42-49.

Koch, S. (1999). *Psychology in human context. Essays in dissidence and reconstruction*. Chicago: The University of Chicago Press.

Köhler, T.& Troester, U. (1991). Changes in the palmar sweat index during mental arithmetic. *Biological Psychology*, 32,143-154.

Kunz, H. (1980) *Leistungsbestimmende Faktoren im Zehnkampf. Eine Längsschnittstudie an Schweizer Spitzeathleten*. Dissertation an der Sozial- und verhaltenswissenschaftlichen Fakultät der Universität Heidelberg.

Lakatos, I. (1977). *The Methodology of Scientific Research Programmes: Philosophical Papers* Volume 1. Cambridge: Cambridge University Press.

Lambert, M. J., Hannöver, W., Nisslmüller, K., Richard, M., & Kordy, H. (2002). Fragebogen zum Ergebnis von Psychotherapie: Zur Reliabilität und Validität der deutschen Übersetzung des Outcome Questionnaire 45.2 (OQ-45.2). *Zeitschrift für Klinische Psychologie und Psychotherapie, 31* (1), 40-46.

Langner, E. & Olbrich, M. (1972). Untersuchungen zur faktoriellen Struktur des I-S-T von Amthauer – mit einem Vergleich der Ergebnisse der Intelligenztests der Eignungs- und Verwendungsprüfung. *Wehrpsychologische Untersuchungen*, *3*, 7 -72..

Lawler, D. A., Ebrahim, S., May, M., & Smith, G. D. (2004. (Mis)use of factor analysis in the study on Insulin Resistance Syndrome. *American Journal of Epidemiology, 159, 11*, 1013-1018.

Lenk. W. (1983) Faktorenanalyse: ein Mythos? Historische und konzeptionelle Untersuchungen zur Faktorenanalyse und Intelligenzforschung. Weinheim: Beltz.

Leventhal, D. S., & Stedman, D. J. (1970). A factor analytic study of the Illinois Test of Psycholinguistic Abilities. *Journal of Clinical Psychology, 26*, 473-476.

Lienert, G. A. (1969). *Testaufbau und Testanalyse*. Dritte, durch einen Anhang über Faktorenanalyse ergänzte Auflage. Weinheim: Beltz.

Linden, M. (1977). Factor analytical study of Olympic decathlon data. *Research Quarterly*, *48,* 562-568.

Lukesch, H. & Kleiter, G. D. (1974). Die Anwendung der Faktorenanaly-se.Darstellung und Kritik der Praxis einer Methode. *Archiv für Psychologie*, *126*, 265-307.

Lumsden, J. (1961). The construction of unidimensional tests. *Psychological Bulletin, 58*, 122-131.

Lykken, D. T. (1991). What's Wrong with Psychology Anyway? In D. Cicchetti and M. Grove (Eds.). *Thinking Clearly about Psychology*. Volume 1: *Matters of Public Interest*. University of Minnesota Press.

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research Monographs, 38* (*1*), 113-139.

MacCallum, R. C., & Austin, J. T. (2000). Application of structural equation modelling in psychological research. *Annual Review of Psychology, 51*, 201-226.

Manning, J. M. (1987). Factor Analysis of Various Anaerobic Power Tests. Paper presented at the National Convention of the American Alliance for Health, Physical Education, Recreation and Dance (Las Vegas, NV, April 13-17, 1987.

Marshall, G. N., Wortman, C. B., Kusulas, J. W., Hervig, L. K., & Vickers, R. R. (1992). Distinguishing optimism from pessimism: Relations to fundamental dimensions of mood and personality. *Journal of Personality and Social Psychology, 62 (6),* 1067-1074).

Marx,A. & Hejj, A. (1989). *Subjektive Strukturen. Ergebnisse aus der Sprach- und Einstellungsforschung*. Göttingen: Hogrefe.

Matson, J. L. & Nieminen, G. S. (1987). Validity of measures of conduct disorder, depression, and anxiety. *Journal of Clinical Child Psychology, 16(2),* 151-157.

Maturana, H. R. (1998). *Biologie der Realität.* Frankfurt: Suhrkamp.

Maxwell, A. E. (1959). Statistical methods in factor analysis. *Psychological Bulletin, 56*, 228-235.

Maxwell, A. E. (1972). Factor analysis: Thomson's sampling theory recalled. *Britisch Journal of Statistical Psychology, 25*, 1-21.

Meili, R. (1969). Die Zwiespältigkeit faktorieller Untersuchungen der Intelligenz-struktur. *Schweizerische Zeitschrift für Psychologie, 16*, 278-294.

Metzger, W. (1941). *Psychologie. Die Entwicklung ihrer Grundannahmen seit der Einführung des Experiments.*Darmstadt: Steinkopf (Letzte Auflage 2002, Wien: Krammer).

Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory and Psychology, 10*, 639-667.

Milhorn, H. T. (1982). Cardiovascular fitness. *American Family Physician, 26 (3),* 163-169.

Mitchell, J. (1997).Quantitative science and the definition of measurement in psychology. *Britisch Journal of Psychology, 88,* 355-383.

Moosbrugger, H. & Hartig, J. (2002). Factor analysis in personality research: Some artifacts and their consequences for psychological assessment. *Psychologische Beiträge 44,* 136-158.

Mulaik, S. (1972). *The foundations of factor analysis.* New York: McGraw Hill.

Narens, L. & Luce, R. D. (1993). Further comments on the « nonrevolution » arising from axiomatic measurement theory. *Psychological Science, 4*, 127-130.

Neuhaus, J. O., & Wrigley, C. F. (1954). The quartimax method: an analytic approach to orthogonal simple structure. *British Journal of Statistical Psychology, 7*, 81-91.

Orlik, P. (1967). Das Dilemma der Faktorenanalyse - Zeichen einer Aufbaukrise in der modernen Psychologie. Sonderheft: Münchener Symposium über Faktorenanalye. *Psychologische Beiträge, 10*, 87-98.

Pawlik, K. (1977). Faktorenanalytische Persönlichkeitsforschung. In G. Strube (Ed.), *Binet und die Folgen. Testverfahren, Differentielle Psychologie, Persönlichkeitsforschung. Bd. 5. Die Psychologie des 20. Jahrhunderts*. Zürich: Kindler.

Peel, E. A. Factorial analysis as a psychological technique. *Nordic Psychology Monographs*, 7-22.

Pett, M. A., Lackey, N. R.,& Sullivan, J. J. (2003). *The use of factor analysis for instrument development in health care research.* Beverly Hills, CA: Sage.

Pöhlmann, R., Kirchner, G., Wohlgefahrt, K.H. (1979) Der psychomotorische Fähigkeitskomplex – seine Kennzeichnung und seine Vervollkommnung. *Theorie und Praxis der Körperkultur, 28,* 898-907.

Quine, W. v. O. (1963). On simple theories of a complex world. *Synthese, 15*, 103-106.

Rescher, N. (1985). *Die Grenzen der Wissenschaft.* Stuttgart: Reclam.

Rasch, D. (1962). Die Faktorenanalyse und ihre Anwendung in der Tierzucht. *Biometrische Zeitschrift*, *4,* 15-39.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12(3),* 287-297.

Rescher, N. (1998). *Complexity. A philosophical overview.*New Brunswick: Transactions Publishers.

Revelle, W. (1983). Factors are fictions, and other comments on individuality theory. *Journal of Personality, 51*, 707-715.

Revenstorf, D. (1976). *Lehrbuch der Faktorenanalyse*. Stuttgart: Kohlhammer.

Revenstorf, D. (1980). *Vom unsinnigen Aufwand*. Stuttgart: Kohlhammer.

Reyburne, H. A., & Rath, M. J. (1949). Simple structure: A critical examination. *The British Journal of psychology, 2*, 125-133.

Reymert, R., & Jöreskog, K. G. (1993). *Applied Factor Analysis in the Natural Sciences*. Cambridge: Cambridge University Press.

Rost, J. (2003, Februar). Zeitgeist und Moden empirischer Analysemethoden [45 Absätze]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [On-line Journal], 4(2). Accessible by: <http://www.qualitative-research.net/fqs-texte/2-03/2-03rost-d.htm>.*

Rozeboom, W. W.(1991).Theory and practice of analytic hyperplane optimization. *Multivariate Behavioral Research, 26(1)*, 179-197.

Rummel, R. J. (1970). *Applied Factor Analysis*. Evanston: Northwestern University.

Ruttkowski, W. V. (1974). *Typologien und Schichtenlehren. Bibliographie des internationalen Schrifttums bis 1970*. Amsterdam: Rodopi.

Sabourin, S., Lussier, Y., Laplante, B., & Wright, J. (1990). Unidimensional and multidimensional models of dyadic adjustment: A hierarchical reconciliation. *Psychological Assessment. A Journal of Consulting and Clinical Psychology, 2*, 333-337.

Sassenscheidt, H. & Buggle, F. (1970). Prädiktorvariablen praktisch-technischen verständnisses. *Diagnostica, 16*, 30-41.

Saunders, D. R. (1962). Transvarimax: Some properties of the ratiomax and equamax criteria for blind orthogonal rotation. *American Psychological Association Meeting.*

Schaper, N., & Baumgart, C. (2002). TKI-Teamklima-Inventar. *Zeitschrift für Personalpsychologie, 1(3*), 54-158.

Schmid, J. & Leiman, J. M. (1957).The development of hierarchical factor solutions. *Psychometrika*, *22*, 53-61.

Schmidt-Atzert, L. & Hommers, W. (1996). IST 70: Weitere Daten – weitere Erkenntnisse? Anmerkungen zu Gerd Heydes Beitrag „Weitere Daten zum IST-70“. *Report Psychologie, 21,* 526-528.

Schmidt-Atzert, L. (1997). Replik zur Rezension des IST 70. *Zeitschrift für Differentielle und Diagnostische Psychologie,* 18, 109-112.

Schmidt-Atzert, L., Hommers, W. & Hess, M. (1995). Der I-S-T 70: Eine Analyse und Neubewertung. *Diagnostica, 41,* 108-130.

Schmitt, M. (2000). Vorschlag zur Vereinfachung des Beck-Depressions-Inventars (BDI). *Diagnostica, 46*(1), 38-46.

Schönemann, P. H. & Borg, I. (1996). Von der Faktorenanalyse zu den Strukturgleichungsmodellen. In Erdfelder, E., Mausfeld, R., Meiser, T., and Rudinger, G. (Eds): *Quantitative Psychologische Methoden - Ein Handbuch in Schlusselbegriffen*, 241-252.

Schönemann, P. H. & Steiger, J. H. (1976). Regression component analysis. *British Journal of Statistical Psychology*, *29,* 175-189).

Schönemann, P. H. & Wang, M. M.(1972). Some new results on factor indeterminacy. *Psychometrica, 37*, 61-91.

Schönemann, P. H. (1981). Measurement, scaling, and factor analysis. In I.Borg (Ed.), *Multidimensional data representation: When and why*. Ann Arbor: Mathesis.

Schönemann, P. H. (1994). Book Review: Louis Guttman on Theory and Methodology: Selected writings. *Applied Psychological Measurement*, *18 (3*), 293-297.

Sechrest, L. B. (2002).Paper without available title presented at the 11th European Conference on Personality. Universität Jena, 21.7.-25.7.2002.

Seitz, W. & Löser, G. (1969). Über die Beziehung von Persönlichkeitsmerkmalen zu Schul- und Intelligenztest-Leistungen bei Gymnasialschülern. *Zeitschrift für experimentelle und angewandte Psychologie,* 16, 651-679.

Singer, W. (2003). *Ein neues Menschenbild? Gespräche über Hirnforschung.* Frankfurt a. M.: Suhrkamp

Smith, P. T. & Jones, K.F.(1975). Some hierarchical scaling methods for confusion matrix analysis. II. Applications to large matrices. *British Journal of Mathematical and Statistical Psychology*, *28,* 30-45.

Snyder, M., & Gangestad, S. (1986.). On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of Personality and Social Psychology, 51*, 125-139.

Sponsel, R. (1998). Axiome und Konstruktionsprinzipien Differentieller Psychologie der Persönlichkeit in der Allgemeinen und Integrativen Psychodiagnostik, Psychologie, Psychopathologie und Psychotherapie. *Internet-Publikation für Allgemeine und Integrative Psychotherapie IP-GIPT. Erlangen: http://www.sgipt.org/gipt/diffpsy/giptpt0.htm.*

Steiger, J. H. & Schönemann, P.H. (1975). A history of factor indeterminacy. *Unpublished manuscript.* Purdue University.

Steiger, J. H. (1994). Factor analysis in the 1980's and the 1990's: Some old debates and some new developments. In I. Borg & P. P. Mohler (Eds.) *Trends and perspectives in empirical social research* (pp. 201-223). New York: De Gruyter.

Stephenson, W. (1956). Methodology of trait analysis. *The British Journal of Psychology, 47*(5-18).

Suppes, P. & Zinnes, J. L. (1963). Basic measurement theory. In: Luce, R. D., Bush, R. R. & Galanter, E. (eds.) *Handbook of Mathematical Psychology.* Vol. I. New York: Wiley.

Szopa,J., Chwała, W., & Ruchlewicz,T.(1998). Investigations on Structure of "Energetic" Motor Abilities and Validityof their Testing. *Antropomotoryka, 17* (Retrieved from the web).

Teipel, D. (1988). *Diagnostik koordinativer Fähigkeiten.* München.

Thalbourne, M. A. (1998). Transliminality: Further correlates and a short measure. *The Journal of the American Society for Psychical Research, 92*, 402-419.

Thompson, J. W. (1962). Meaningful and unmeaningful rotation of factors. *Psychological Bulletin, 59 (3)*, 211-223.

Thompson, J. W. (1963). Bi-polar and uni-directional scales. *British Journal of Psychology, 54*, 15-24.

Thurstone, L. L. (1934). The Vectors of the Mind. *Psychological Review, 41*, 1-32.

Thurstone, L. L. (1935). *The Vectors of the Mind.* Chicago: The University of Chicago Press.

Thurstone, L. L. (1947). *Multiple factor analysis. A development and expansion of The Vectors of the Mind.* Chicago: The University of Chicago Press.

Tinsley, H. E.A& Tinsley, D. J. (1987). Uses of factor analysis in Counseling Psychology research. *Journal of Consulting Psychology, 14,* 414-424.

Tversky, A. (1977). Features of similarity. *Psychological Revue, 84,* 6,327-348

Überla, K. (1971). *Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler.* 2. Auflage, Berlin: Springer. Erste Auflage 1968, letzte 1977.

Undheim, J. O., & Gustavsson, J.-E. (1987). The hierarchical organization of cognitive abilities. Restoring general intelligence through the use of structural relations (LISREL). *Multivariate Behavioral Research Monographs, 22*, 149-171.

Velicer, W. F. & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25(1*), 1-28.

Velicer, W. F. (1977). An empirical comparison of the similarity of principal component, image, and factor patterens. *Journal of Multivariate Behavioral Research, 12*, 3-22.

Vittadini, G., 1989. Indeterminacy problem in the LISREL model. *Multivariate Behavioral Research, 24 (4),* 397-414.

Vukovich, A. (1967). Faktorielle Typenbestimmung. Sonderheft: Münchener Symposium über Faktorenanalye. *Psychologische Beiträge, 10*, 81-86.

Waldmann,M. R., Renkl, A. & Gruber, A. (2003).Das Dreieck von Begabung,Wissen und Lernen.. In W. Schneider & M. Knopf. (Hrsg.) *Entwicklung, Lehren und Lernen.*Göttingen: Hogrefe.

Warburton, F. W. (1963). Analytic methods of factor rotation. *The British Journal of Statistical Psychology, 16*, 165-174.

Watson, D., & Clark, L. A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Weber, E. (1978). *Einführung in die Faktorenanalyse*: *mit einem Anhang zur Matrizenrechnung*. - 1. Aufl. 1974. Stuttgart : Fischer

Weinert, F. E. (Hrsg.) (1996). *Psychologie des Lernens und der Instruktion*. Göttingen: Hogrefe.

Wellek, A. (1959). *Die Polarität im Aufbau des Charakters*. Bern: Francke.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*, 395-412.

Wiley, D. E. (1967). Latent partition analysis. *Psychometrica, 32(2),* 183-193.

Witte, W. (1974). Untersuchungen zur Behinderung des Denkens durch Anschauung. *Psychologische Beiträge, 16*, 277-287.

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. New York: State University of New York Press.

Zarnowski, F. (1989) *The decathlon. A colorful history of track and field's most challenging events*. Champaign, Ill.: Leisure Press.

Zimmermann, W. S. (1953). A note on the recognition and interpretation of composite factors. *Psychological Bulletin*, 50, 387-389.

Ein fataler Fehler der konventionellen Psychometrie wird in dieser Monographie aufgedeckt. Das Prinzip der Einfachstruktur, an dem sich seit den Anfängen der Faktorenanalyse alle Varianten der Rotation extrahierter Faktoren ausgerichtet haben, geht an der empirischen Realität vorbei. Das „Simple Structure"-Prinzip zwingt die Variablen einer multivariaten Untersuchung in ein monokausales Korsett und hindert sie daran, die tatsächlich vorliegende Komplexität latenter Komponenten zu entfalten. Dies ist der Grund, warum faktorenanalytische Ergebnisse durchweg theoretisch fruchtlos blieben und eher Unbehagen als Zufriedenheit mit sich brachten. Mit den später entwickelten konfirmatorischen Verfahren, die am realitätsfremden Ideal der Einfachstruktur festhielten, wurde keine Wende erzielt. Eine Wende aber stellt sich ein, sobald man sich dazu entschließt, bei der Rotation von Faktoren das scheinbar unverzichtbare Prinzip der Einfachstruktur aufzugeben und mehrfaktorielle Komplexität der Variablen unbeschränkt zuzulassen. Dazu dient in der statistischen Praxis das vom Verfasser entwickelte Rotationsverfahren Varimin. Die Ergebnisse von Varimin-Analysen, die in diesem Buch beschrieben werden, übertreffen die Ergebnisse von Einfachstruktur-Analysen (Varimax) an Deutbarkeit, Stabilität, und – scheinbar paradoxerweise – sogar auch an Einfachheit.



ISBN: 978-3-86395-013-2

Universitätsverlag Göttingen